METRON

RIVISTA INTERNAZIONALE DI STATISTICA — REVUE INTERNATIONALE DE STATISTIQUE

INTERNATIONAL REVIEW OF STATISTICS — INTERNATIONALE STATISTISCHE RUNDSCHAU

DIRETTORE PROPRIETARIO — DIRECTEUR ET PROPRIÉTAIRE
EDITOR AND PROPRIETOR — HERAUSGEBER UND EIGENTHÜMER

Dott. Corrado Gini, *prof. ord. di Statistica nella R. Università di Padova* (Italia).

Vol. II. N. 4. **1 - VI - 1923**

SOMMARIO — SOMMAIRE — CONTENTS — INHALT

A. Guldberg, *Zur Theorie der Korrelation.*

A. A. Tschuprow, *On the mathematical expectation of the moments of frequency distributions in the case of correlated observations* (cont.).

J. W. Bispham, *An experimental determination of the distribution of the partial correlation coefficient in samples of thirty.*

R. Pearl, *The Interrelation of the Biometric and Experimental Methods of acquiring Knowledge; with special reference to the Problem of the Duration of Life.*

L. Colomba, *La statistica e le scienze naturali.*

J. Bokalders, *Lettlands Agrarproblem.*

M. Boldrini, *La décroissance sénile chez l'homme et chez la femme.*

G. Tagliacarne, *Contributi e comportamenti delle regioni d'Italia in guerra.*

A. MacDonald, *Death Psycology of Historical Personages.*

FERRARA (ITALIA)

CASA EDITRICE TADDEI

45 Via de' Romei

Articoli giunti alla Rivista che verranno pubblicati nei prossimi numeri.
(Secondo l'ordine d'arrivo)

Artikel die an die Rundschau angelangt sind und welche in den nachfolgenden Nummern erscheinen werden.
(Nach der Reihenfolge des Eingangs)

Articles reçus par la Revue et à paraître prochainement.
(D'après la date de réception)

Articles received by the Review which will be published in future issues.
(According to date of receipt)

A. Guldberg, *Ueber Markoff's Ungleichung.*

H. Ziemann, *Beitrag zur Bevölkerungsfrage der farbigen Rassen.*

R. Pearl and **L. J. Reed,** *On the mathematical theory of population growth.*

M. Boldrini, *Tables des reductions annuelles des contingents militaires.*

F. v. Fellner, *Die Verteilung des Volksvermögens und Volkseinkommens der Länder der Ungarischen Heiligen Krone zwischen dem heutigen Ungarn und den Successions - Staaten.*

F. Savorgnan, *Nuzialità e fecondità delle Case Sovrane d'Europa.*

C. Gini, *Quelques chiffres sur la richesse et les revenus nationaux de quinze Etats.*

C. Gini, *Sulle leggi della frequenza e delle combinazioni sessuali dei parti plurimi.*

W. Feld, *Internationale Bibliographie der Statistik der Kindersterblichkeit.*

C. Gini e **L. De Berardinis,** *Sulle vaccinazioni antitifiche nell'esercito italiano durante la guerra.*

C. Gini e **M. Boldrini,** *Il centro della popolazione italiana.*

H. W. Methorst, *Prédominance des naissances masculines.*

Alf Guldberg

Zur Theorie der Korrelation

Sehr verschiedene Meinungen scheinen über die Begriffe Korrelation und Korrelations-Koefficient zu herrschen. Um dies zu beleuchten erlaube ich mir zu citieren.

Professor J. C. Kapteyn schreibt (1):

« There is no doubt, what is meant by perfect correlation — correlation-coefficient $r = 1.00$ — between the two varying quantities x and y. It means that any variation of x is accompanied by a definite variation of y. It implies that y is a pure function of x. Neither is there a doubt about what is meant by $r = 0.00$, that is, absolute absence of correlation. It means that the variation of x and y are *perfectly independent of each other* ».

Professor H. L. Rietz schreibt (2):

« The recent papers of Reed and Harris in these Publication have brought to my mind some simple cases in which a correlation coefficient is *zero* although the *two variables* are *mathematically functions* of each other represented by certain simple types of continuous curves ».

Professor C. V. L. Charlier schreibt (3):

« Ist der Korrelations-koefficient r gleich null, so bedeutet dies, dass die betrachteten statistischen Erscheinungen von einander unabhängig sind und nichts mit einander zu thun haben ».

---

(1) « Monthly Notices of Royal Astronomical Society » v. 72 n 6, April 1912, p. 518.

(2) « Quarterly Publications of the American Statistical Association » September, 1919, p. 472.

(3) C. V. L. Charlier - *Vorlesungen über die Grundzüge der mathematischen Statistik*, Hamburg, 1920 p. 90.

Professor G. UDNY YULE schreibt (4):

« It should be noted that, while r is zero if the variables are independent, the converse is not necessarily true ».

Die Anschaungen sind, wie man sieht, direct entgegengesetzt. Ich denke daher, dass die folgende Darstellung der Korrelationstheorie von einiger Interesse sein könnte, da keine dergleichen Gegensätze vorkommen.

Die erste Definition von dem Begriffe Korrelation stammt, so viel ich weiss, von Sir FRANCIS GALTON.

Er schreibt (5):

« Two variable organs are said to be co-related, when the variation of the one is accompanied on the average by more or less variation of the other, and in the same direction ».

Die Untersuchungen über Korrelation wurde von Professor KARL PEARSON fortgesetzt. Professor PEARSON gab folgende Definition (6):

« Two organs in the same individual or in a connected pair of individuals are said to be correlated, when a series of the first organ of a definite size being selected, the mean of the sizes of the corresponding second organs is found to be a function of the size of the selected first organ. If the mean is independent of this size, the organs are said to be non - correlated ».

Der Sinn dieser Definition ist, denke ich, folgender.

Man hat eine Menge — eine Verteilung — von gleichartigen Objecte, die in Bezug auf zwei veränderliche zahlenmässige darstelbare Merkmale karakterizirt sind. Die Merkmale (*the organs*) nennen wir Variable oder Argumente. Wir bezeichnen die Variable mit x and y.

Wir betrachten nun diejenigen Objecte unserer Verteilung, wo die eine der Variablen, z. b. x, einen vorgeschriebenen Wert x_0 hat. Wir bestimmen den *Mittelwert* der anderen Variablen y aller dieser Objecte. Wir bezeichnen diesen Mittelwert der Variablen y mit $\bar{y}_0$. Wir betrachten ferner diejenigen Objecte unserer Verteilung, wo die Variable x einen anderen vorgeschriebenen Wert x_1 hat.

---

(4) G. UDNY YULE - *An Introduction to the Theory of Statistics*, London 1919, p. 174.

(5) « Proc. Roy. Soc. » v. 45, 1889, p. 135.

(6) « Phil. Trans. Roy. Soc. » A. 187; 1895, p. 233.

Wir bestimmen den *Mittelwert* der anderen Variablen y aller dieser Objecte. Wir bezeichnen diesen Mittelwert der Variablen y mit $\bar{y}_1$ u. s. w. Wenn die Verteilung unserer Objecte so beschaffen ist, dass wenn x_0 von x_1 verschieden ist, dann auch $\bar{y}_0$ von $\bar{y}_1$ verschieden ist, oder allgemein gesprochen, wenn der Mittelwert der Variablen y für einen vorgeschriebenen Wert von x variirt, wenn x variirt, so sind die Variablen nach Pearson korreliert. Die Korrelation der Variablen hängt also von der *Verteilung* unserer Objecte ab. Die Variablen x und y sind mathematisch unabhängige Variablen, die unsere Objecte karakterisieren.

Ein Beispiel (1) einer Verteilung eines Objects karakterisiert bei zwei unabhängigen Merkmalen ist die folgende so gennante Korrelationstafel. Die Tafel enthält 5.317.000 *Ehen*, karakterisirt bei den Alterstufen x und y für Mann und Frau. Die Zahlen sind 1000.

		Alter der Frau															Summe
		15-	20-	25-	30-	35-	40-	45-	50-	55-	60-	65-	70-	75-	80-	85-	
Alter der Mann	15-	2	2														4
	20-	16	173	46	4	1											204
	25-	4	185	402	84	10	2	1									688
	30-	1	41	265	411	84	12	2	1								817
	35-		9	69	251	369	80	12	2	1							793
	40-		3	17	71	219	309	66	12	2	1						700
	45-		1	6	20	66	178	252	59	10	2	1					595
	50-			2	8	19	57	146	195	44	10	2					483
	55-			1	3	8	18	46	110	141	35	6	1				369
	60-				1	3	8	16	39	81	101	23	4	1			277
	65-				1	1	3	6	11	26	53	58	13	2	1		175
	70-					1	1	2	5	8	18	31	31	6	1		104
	75-						1	1	1	3	5	10	14	12	2		50
	80-								1	1	1	1	4	5	3	1	18
	85-											1	1	1	1		4
Summe		23	414	808	854	781	669	550	437	317	226	134	65	27	8	1	5317

Tabellen dieser Art bilden die statistische Grundlage unserer Untersuchung. Das allgemeine Problem ist die Bestimmung des ana-

(1) G. U. Yule, l. c. p. 159.

lytischen Ausdruches $f(x, y)$ des mathematischen Verteilungsgesetzes unserer Objecte, die von den zwei unabhängigen Variablen x und y abhängen, aus den gegebenen statistischen Erhebungen. Dies allgemeine Problem habe ich an einer anderen Stelle (1) behandelt. Der Zweck dieser Linien ist eine Darstellung des Korrelations-Problems an ein specielles Verteilungsgesetz.

Wenn wir ein Object z betrachten, das nur von einer einzigen Variablen x abhängt, wissen wir, dass viele solcher Objecte das Gaussische Verteilungsgesetz $z = Ce^{-h^2x^2}$ folgen.

Betrachten wir ein Object, das von zwei unabhängigen Variablen x und y abhängt, zeigt es sich, dass oft solche Objecte das generalisierte Gaussische Verteilungsgesetz, erst abgeleitet von Bravais $Z = C e^{-(ax^2+2bxy+cy^2)}$ folgen.

Das Bravaische Verteilungsgesetz wird im allgemeinen, wenn wir die Konstanten bestimmt haben, in der folgenden Form (2) geschrieben:

$$z = \frac{N}{2\pi\sigma_x.\sigma_y\sqrt{1-r^2}}\, l^{-\frac{1}{2(1-r^2)}\left[\frac{x^2}{\sigma x^2} - \frac{2r}{\sigma x\, \sigma y} x y + \frac{y^2}{\sigma y^2}\right]}$$

wo Origo in dem Mittelwert der Variablen x und y liegt, N die totale Anzahl der Objecte, $\sigma_x^2, \sigma_y^2, M_{(x,y)}$ die Mittelwerter von $x^2, y^2, x.y$ und $r = \frac{M(x,y)}{\sigma_x.\sigma_y}$ sind; r ist der Korrelationskoefficient.

Ist das Verteilunggesetz eines Objectes:

$$z = f(x,y)$$

und bezeichnen wir mit $\overline{y}_{x_0}$ den Mittelwert von der Variable y für einen vorgeschriebenen Werth x_0 der Variable x, so ist

$$\overline{y}_{x_0} = \frac{\int_{-\infty}^{\infty} y f(x_0,y) dy}{\int_{-\infty}^{\infty} f(x_0,y)\, dy}$$

---

(1) « Comptes Rendus du Congrès International des Mathématiciens », Strasbourg, 1921.

(2) Vgl. z. b. Elderton, *Frequency curves and correlation*, p. 112.

und analog, wenn $\overline{x}_y$, der Mittelwert der Variablen x für einen vorgeschriebenen Werth y_0 der Variable y ist

$$\overline{x}_{y_0} = \frac{\int_{-\infty}^{\infty} x f(x, y_0) dx}{\int_{-\infty}^{\infty} f(x, y_0) dx.}$$

Das fundamentale Problem der Korrelation ist, die Werte $\overline{y}_{x_0}$ und $\overline{x}_{y_0}$ als Funktionen von x_0 respective y_0 zu bestimmen.

Folgt unser Object das Bravaische Verteilungsgesetz, oder hat man, wie man sagt, normale Korrelation, so findet man nach einer kleinen Rechnung:

$$\overline{y}_{x_0} = r \frac{\sigma_y}{\sigma_x} \cdot x_0$$

Variiren wir x_0, so erhalten wir die Gleichung der *Regressionslinie* von y auf x.

Ganz analog finden wir:

$$\overline{x}_{y_0} = r \frac{\sigma_x}{\sigma_y} \cdot y_0$$

Variiren wir x_0 so erhalten wir die Gleichung der *Regressionslinie* von x auf y. Wenn unser Object das Bravaische Verteilungsgesetz folgt, sind die Regressionslinien *gerade* Linien. Die Werte der Variablen x und y für welche unser Object am häufigsten eintrifft, werden die *wahrscheinlichsten* Werte der Variablen genannt. Teilt man dereinen der variablen z. b. x einen vorgeschriebenen Wert x_0 an, so nennt man den Wert der anderen Variablen, hier y, für welche unser Object am häufigsten erscheint, den wahrscheinlichsten Wert dieser Variable. Das Problem der Bestimmung der wahrscheinlichsten Werte der Variablen, wenn das Verteilungsgesetz $z = f\ (x, y)$ unserer Objecte bekannt ist, ist ein einfaches Maximum-Problem.

Um den wahrscheinlichsten Wert von y für $x = x_0$ zu finden, haben wir y aus der Gleichung

$$\frac{\delta f_{(y, x_0)}}{\delta_y} = o$$

zu bestimmen.

Folgt unser Object das Bravaische Verteilungsgesetz, findet man

$$y = r \cdot \frac{\sigma_y}{\sigma_x} \cdot x_0$$

Wir finden also den selben Ausdruck für den wahrscheinlichsten Wert der Variable y für $x = x_0$ als für den *Mittelwert* der Variable y für $x = x_0$.

Ganz analog finden wir für den wahrscheinlichsten Wert von x für einen vorgeschriebenen Werte $y = y_0$

$$x = r \frac{\sigma_x}{\sigma_x} \cdot y_0$$

Variiren wir in den zwei letzten Gleichungen x_0 und y_0, erhalten wir die Gleichungen der *Regressionslinien* für die wahrscheinlichsten Werte respective von y auf x und von x auf y.

Folgt unser Object das Bravaische Verteilungsgesetz, fallen also die Regressionslinien für die Mittelwerte und die wahrscheinlichsten Werte zusammen.

Bezeichnen wir mit u den Winkel, den die zwei Regressionslinien mit einander bilden, hat man

$$tg_u = \frac{1-r^2}{r} \frac{\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}.$$

Von dem analytischen Ausdruck des Bravaischen Verteilungsgesetzes, sieht man dass man hat $1 - r^2 > 0$ d. h. $r^2 < 1$. Nähert r sich zum 1, nähern die zwei Regressionslinien sich zu einander. Ist $r = 0$, stehen die zwei Regressionslinien senk recht auf einander. Ihre Gleichungen sind da:

$$\bar{y}_x = 0 \text{ und } \bar{x}_y = 0$$

d. h. die Mittelwerte der Variablen y respective die Mittelwerte der Variablen x sind unabhängige von irgend einem vorgeschriebenen Wert x_0 von respective y_0 von y. Wir haben kein Korrelation.

Bei dieser Darstellnng der Korrelation sind die Variablen x und y unabhängig entweder Korrelation existirt oder nicht ($r \gtrless 0$). $r = 0$). Wir sehen anch direct den Sinn des Korrelationskoefficient, wenn unser Object das Bravaische Verteilungsgesetz folgt.

Die Frage ist vielfach diskutirt, welche Bedeutung der Korrelations-Koefficient hat, wenn unser Object ein willkürliches Verteilungsgesetz folgt.

In einer bekannten Abhandlung hat Mr. YULE diese Frage behandelt. Er schliesst die Abhandlnng mit folgender Wörte (1):

« We can now see, that the use of normal regression formulæ is quite legitimate in all cases, so lang as the necessary limitations

(1) « Proc. Roy. Soc. London », v. 60, p. 489.

of interpretations are recognized. Bravais r always remains a coefficient of correlation ».

Mr. YULE geht in seiner Abhandlung von der Voraussetzung aus dass die Mittelwerte der Variable y respective x für einen bestimmten Wert x_0 von x respective y_0 von y *annährungsweise* wenn x_0 respective y_0 variiren, auf *gerade* Linien liegen. Unter diese Voraussetzung bestimmt Mr. YULE 2 gerade Linien, nach der Methode der kleinsten Quadrate, die die 2 Mittelwertscurven approximieren. Die Gleichungen dieser so definirte gerade Linien haben dieselbe Form wie die früher, unter Voraussetzung des Bravaissche Verteilungsgesetz, gefundene Regressionlinien. Nur unter diese specielle Voraussetzung der Mittelwertcurven gilt also Mr. *Yulle's* Aussage.

Uebrigens ist es unmittelbar klar, dass es unmöglich ist eine willkürliche Verteilungsgesetz durch eine einzige Konstante zu karakterisieren.

Der Begriff Korrelation, der wir bis jetzt betrachtet haben, ist unlösbar mit einem *Verteilungsgesetz* verbunden.

Der Begriff Korrelation wird aber anch bei einem *ganz anderem* statistischen Problem gebraucht. Ich citire das Problem nach M. MARCEL LENOIR (1):

« Ayant deux courbes representant deux séries numériques relatives à deux phénomènes, désignons par x_1 et x_2 les écarts à leurs valeurs moyennes dans tout l'intervalle considéré de deux nombres correspondants . . . .

. . . .

Si les deux courbes considérées sont tout à fait semblabes, les écarts x_1, x_2 sont dans un rapport constant, et l'on a toujours, quel que soit le couple considéré $x_1\,x_2$, $x_1 = b_2\,x_2$, ou $x_2 = b_1\,x_1$, et, en ce cas $b_1\,b_2 = 1$. En tout cas, même si les deux courbes ne sont pas tout à fait semblables, on pourra toujours déterminer un certain ceefficient b_2 tel que l'equation $x_1 = b_2\,x_2$ donne pour x_1, si l'on remplace x_2 par ses valeurs réelles, les meilleurs valeurs possibles. La méthode dite: méthode des moindres carrées, permet de calculer b_2 de manière que que la somme des carrées des différences entre les vraies valeurs de x_1 et les valeurs calculées en prenant $x_1 = b_2\,x_2$, soit la plus petite possible. L'equation $x_1 = b_2\,x_2$ qui permet ainsi de

(1) M. LENOIR: *Etudes sur la formation et le mouvement des prix*, p. 69 Das Problem behandelt auch C. V. L. CHARLIER, l. c. p. 86; G. U. YULE: *Theory of Statistics*, London, p. 197; S. D. WICKSELL: *Elementen ar Statistikens Teori*, p. 17 und p. 76.

calculer les meilleurs valeurs approchées de x_1 est dite *equation de régression*. Le valeur de b_2 est donnée par la formule $b_2 = \frac{\Sigma x_1 x_2^2}{\Sigma x_2}$. On peut calculer de même une équation de regression qui donnera les meilleurs valeurs approchées de x_2 correspondant aux $x_1, x_2 = b_1, x_1$, et içi la valeur de b_1 est $b_1 = \frac{\Sigma x_1 x_2}{\Sigma x_1^2}$.

Si les deux équations de régression étaient rigoureusement exactes, on aurait, comme nous l'avons dit, $b_1 b_2 = 1$. Ce produit différe peu de 1 si les deux équations donnent des valeurs de x_1 et x_2 peu differents des vraies valeurs. On peut donc considérer le coefficient $r = \sqrt{b_1 b_2}$ comme mesurant l'exactitude des équations de régression, et par conséquent la rassemblance des deux courbes. Le coefficient

$$r_{12} = \sqrt{b_1 b_2} = \frac{\Sigma x_1 x_2}{\sqrt{\Sigma x_1^2 . \Sigma x_2^2}}$$

n'est pas autre chose que le coefficient de corrélation, au signe prés».

Wir sehen, dass das hier betrachtete Problem ganz verschieden von Professors PEARSONS Korrelations-Problem ist.

Wir bezeichnen mit

$$y_1, y_2 \cdots y_n$$

$$x_1, x_2, \ldots x_n$$

die Abweichungen von ihren respectiven Mittelwerten zweier statistichen Reihen, wo x_i und y_i gleichzeitig beobachtet sind.

Wir können da in unendlich vieler Weise eine Funktion $y = F(x)$ angeben, die für die Werte $x_1, x_2, \ldots x_n$ die Werte $y_1, y_2, \ldots y_n$ annehmen. LAGRANGE's interpolation formula gibt z. b. eine *ganze* Function $(n\text{-}1)^{ste}$ Grades dieser Art:

$$y = \frac{y_1}{(x-x_1) f'_{(x_1)}} \cdot f_{(x)} + \frac{y_2}{(x-x_2) f'_{(x_2)}} f_{(x)} + \ldots + \frac{y_n}{(x-x_n) f'(x_n)} f_{(x)}.$$

wo

$$f_{(x)} = (x-x_1)(x-x_2) \ldots (x-x_n)$$

Wir haben also in dieser Weise eine *funktionale* Relation zwischen die beobachteten Variablen x un y.

Wir können natürlich immer eine gerade Linie $y = bx$ zu der Curve, deren Gleichung die gefundene Funktion $y = F_{(x)}$ ist, nach der Methode der kleinsten Quadrate approximiren. Wir finden da

$b_1 = \frac{\Sigma xy}{\Sigma x^2}$. Setzen wir:

$$r = \frac{\Sigma xy}{n.\sigma_x.\sigma_y}\ ;\ \sigma_x^2 = \frac{\Sigma x^2}{n}\ ,\ \sigma_y^2 = \frac{\Sigma y^2}{n}$$

lautet die Gleichung dieser gerade Linie

$$y = r\frac{\sigma_y}{\sigma_x}\cdot x$$

Diese gerade Linie hat selbst verständlich nicht die Bedeutung der Regressionslinie von y auf x bei unseren frühere Korrelationsproblem.

Wir können in diesem Fall sehr gut $r = 0$ finden (d. h. dass die approximirte gerade Linie die *x-a x e* ist) wo eine funktionale Verbindung zwischen y und x besteht (cfr. H. L. Rietz l. c.)

Die zwei hier genannten Probleme werden leider meistens nicht aus einander gehalten. Ich schliesse mich vollständig Professor Corrado Gini an:(1) «La teoria della misura delle relazioni tra i caratteri costituisce indubbiamente uno dei campi della metodologia statistica maggiormente suscettibili di progresso. Condizione necessaria per tale progresso è che vengano debitamente distinti e precisati i varii aspetti sotto cui possono essere considerate le relazioni tra i caratteri ».

(1) «Atti del R. Istituto Veneto di Scienze», 1915, t. 124, p. 609.

AL. A. TSCHUPROW

On the mathematical expectation of the moments of frequency distributions in the case of correlated observations

CHAPTER IV

§ I

1) Let us distribute the N observations in S groups, the first of which contains the first n_1 observations, the second of which contains the following n_2 observations and so on $(n_1 + n_2 + \ldots + n_s = N)$. Denoting by $X'_{i,j}$ the value taken by the corresponding variable in the j-th experiment of the i-th group, put:

$$\frac{1}{n_i}\sum_{j=1}^{n_i} X'_{i,j} = Z'_i \quad , \quad \frac{1}{S}\sum_{i=1}^{S} Z'_i = Z_{(s)}$$

$$\boldsymbol{E}\left[Z'_i\right]^r = M_r^{(i)} \quad , \quad \boldsymbol{E}\left[Z'_i - M_1^{(i)}\right]^r = \underline{M}_r^{(i)}$$

$$\boldsymbol{E}\left[Z'_{i_1}\right]^{r_1}\left[Z'_{i_2}\right]^{r_2}\cdots\left[Z'_{i_l}\right]^{r_l} = M_{r_1,r_2,\ldots,r_l}^{(i_1,i_2,\ldots,i_l)}$$

$$\boldsymbol{E}\left[Z'_{i_1} - M_1^{(i_1)}\right]^{r_1}\left[Z'_{i_2} - M_1^{(i_2)}\right]^{r_2}\cdots\left[Z'_{i_l} - M_1^{(i_l)}\right]^{r_l} = \underline{M}_{r_1,r_2,\ldots,r_l}^{(i_1,i_2,\ldots,i_l)}$$

$$\frac{1}{S^{[-l]}}\,\boldsymbol{S}_l\, M_{r_1,r_2,\cdots,r_l}^{(i_1,i_2,\cdots,i_l)} = M_{[r_1,r_2,\cdots,r_l;S]}$$

$$\frac{1}{S^{[-l]}}\,\boldsymbol{S}_l\, \underline{M}_{r_1,r_2,\cdots,r_l}^{(i_1,i_2,\cdots,i_l)} = \underline{M}_{[r_1,r_2,\cdots,r_l;S]},$$

where $\boldsymbol{S}_l$ denotes, as above (chapter I, § I) a l-fold sum extended to all unequal values of $i_1, i_2, \ldots, i_l$. Let us put further:

$$\boldsymbol{E}\left[X'_{i,j}\right]^r = m_r^{(i \div j)}$$

$$\boldsymbol{E}\left[X'_{i,j_1}\right]^{r_1}\left[X'_{i,j_2}\right]^{r_2} = m_{r_1,r_2}^{(i,i\div j_1,j_2)}$$

$$\boldsymbol{E}\left[X'_{i_1,j}\right]^{r_1}\left[X'_{i_2f}\right]^{r_2} = m_{r_1,r_2}^{(i_1,i_2\div j,f)} \quad \text{and so on.}$$

$$\boldsymbol{E}\left[X'_{i,j} - m_1^{(i\div j)}\right]^{r} = \mu_r^{(i\div j)}$$

$$\boldsymbol{E}\left[X'_{i,j_1} - m_1^{(i\div j_1)}\right]^{r_1}\left[X'_{i,j_2} - m_1^{(i\div j_2)}\right]^{r_2} = \mu_{r_1,r_2}^{(i,i\div j_1,j_2)}$$

$$\boldsymbol{E}\left[X'_{i_1,j} - m_1^{(i_1\div j)}\right]^{r_1}\left[X'_{i_2,f} - m_1^{(i_2\div f)}\right]^{r_2} = \mu_{r_1,r_2}^{(i_1,i_2\div j,f)} \quad \text{and so on}$$

In conformity to the definitions, we have:

$$\frac{1}{N}\sum_{i=1}^{S} n_i\, Z'_i = X_{(N)}$$

$$M_1^{(i)} = \boldsymbol{E}\, Z'_i = \frac{1}{n_i}\sum_{j=1}^{n_i} m_1^{(i\div j)}$$

$$M_2^{(i)} = \frac{1}{(n_i)^2}\sum_{j=1}^{n_i} m_2^{(i\div j)} + \frac{1}{(n_i)^2}\sum_{j_1=1}^{n_i}\sum_{j_2\neq j_1} m_{1,1}^{(i,i\div j_1,j_2)}$$

$$M_{1,1}^{(i_1,i_2)} = \frac{1}{n_{i_1}\, n_{i_2}}\sum_{j=1}^{n_{i_1}}\sum_{f=1}^{n_{i_2}} m_{1,1}^{(i_1,i_2\div j,f)} \quad \text{and so on}$$

$$\underline{M}_1^{(i)} = 0$$

$$\underline{M}_2^{(i)} = \frac{1}{(n_i)^2}\sum_{j=1}^{n_i} \mu_2^{(i\div j)} + \frac{1}{(n_i)^2}\sum_{j_1=1}^{n_i}\sum_{j_2\neq j_1} \mu_{1,1}^{(i,i\div j_1,j_2)}$$

$$\underline{M}_{1,1}^{(i_1,i_2)} = \frac{1}{n_{i_1}\, n_{i_2}}\sum_{j=1}^{n_{i_1}}\sum_{f=1}^{n_{i_2}} \mu_{1,1}^{(i_1,i_2\div j,f)} \quad \text{and so on}$$

2) If all quantities m_r, all quantities m_{r_1,r_2} and so on are equal, we find:

$$M_1^{(i)} = m_1$$

$$M_2^{(i)} = \frac{1}{n_i} m_2 + \frac{n_i - 1}{n_i} m_{1,1}$$

$$M_{1,1}^{(i_1,i_2)} = m_{1,1}$$

$$M_3^{(i)} = \frac{1}{(n_i)^2} m_3 + 3\,\frac{n_i - 1}{(n_i)^2} m_{2,1} + \frac{(n_i - 1)(n_i - 2)}{(n_i)^2} m_{1,1,1}$$

$$M_{2,1}^{(i_1,i_2)} = \frac{1}{n_{i_1}} m_{2,1} + \frac{n_{i_1}-1}{n_{i_1}} m_{1,1,1}$$

$$M_{1,1,1}^{(i_1,i_2,i_3)} = m_{1,1,1}$$

$$M_4^{(i)} = \frac{1}{(n_i)^3} m_4 + \frac{4(n_i-1)}{(n_i)^3} m_{3,1} + \frac{3(n_i-1)}{(n_i)^3} m_{2,2} +$$

$$+ \frac{6(n_i-1)(n_i-2)}{(n_i)^3} m_{2,1,1} + \frac{(n_i-1)(n_i-2)(n_i-3)}{(n_i)^3} m_{1,1,1,1}$$

$$M_{3,1}^{(i_1,i_2)} = \frac{1}{(n_{i_1})^2} m_{3,1} + \frac{3(n_{i_1}-1)}{(n_{i_1})^2} m_{2,1,1} + \frac{(n_{i_1}-1)(n_{i_1}-2)}{(n_{i_1})^2} m_{1,1,1,1}$$

$$M_{2,2}^{(i_1,i_2)} = \frac{1}{n_{i_1} n_{i_2}} m_{2,2} + \frac{n_{i_1}+n_{i_2}-2}{n_{i_1} n_{i_2}} m_{2,1,1} + \frac{(n_{i_1}-1)(n_{i_2}-1)}{n_{i_1} n_{i_2}} m_{1,1,1,1}$$

$$M_{2,1,1}^{(i_1,i_2,i_3)} = \frac{1}{n_{i_1}} m_{2,1,1} + \frac{n_{i_1}-1}{n_{i_1}} m_{1,1,1,1}$$

$$M_{1,1,1,1}^{(i_1,i_2,i_3,i_4)} = m_{1,1,1,1} \quad \text{and so on.}$$

Hence:

$$M_{[1;S]} = m_1$$

$$M_{[2;S]} = m_{1,1} + \left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{n_i}\right][m_2 - m_{1,1}]$$

$$M_{[1,1;S]} = m_{1,1}$$

$$M_{[3;S]} = m_{1,1,1} + 3\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{n_i}\right][m_{2,1} - m_{1,1,1}] +$$

$$+ \left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{(n_i)^2}\right][m_3 - 3\,m_{2,1} + 2\,m_{1,1,1}]$$

$$M_{[2,1;S]} = m_{1,1,1} + \left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{n_i}\right][m_{2,1} - m_{1,1,1}]$$

$$M_{[1,1,1;S]} = m_{1,1,1}$$

$$M_{[4;S]} = m_{1,1,1,1} + 6\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{n_i}\right][m_{2,1,1} - m_{1,1,1,1}] +$$

$$+ \left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{(n_i^2)}\right][3\,m_{2,2} + 4\,m_{3,1} - 18\,m_{2,1,1} + 11\,m_{1,1,1,1}] +$$

$$+\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{(n_i)^3}\right][m_4-4\,m_{3,1}-3\,m_{2,2}+12\,m_{2,1,1}-6\,m_{1,1,1,1}]$$

$$M_{[3,1;S]}=m_{1,1,1,1}+3\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{n_i}\right][m_{2,1,1}-m_{1,1,1,1}]+$$

$$+\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{(n_i)^2}\right][m_{3,1}-3\,m_{2,1,1}+2\,m_{1,1,1,1}]$$

$$M_{[2,2;S]}=m_{1,1,1,1}+2\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{n_i}\right][m_{2,1,1}-m_{1,1,1,1}]+$$

$$+\frac{1}{S\,(S\text{-}1)}\left[\left(\sum_{i=1}^{S}\frac{1}{n_i}\right)^2-\sum_{i=1}^{S}\frac{1}{(n_i)^2}\right][m_{2,2}-2\,m_{2,1,1}+m_{1,1,1,1}]$$

$$M_{[2,1,1;S]}=m_{1,1,1,1}+\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{n_i}\right][m_{2,1,1}-m_{1,1,1,1}]$$

$M_{[1,1,1,1;s]}=m_{1,1,1,1}$ and so on

Substituting for the quantities m the corresponding quantities μ, we obtain $\underline{M}_{r,}^{(i)}$ $\underline{M}_{r_1,r_2}^{(i_1,i_2)}$ an so on, $\underline{M}_{[r;S]}\underline{M}_{[r_1,r_2;S]}$ an so on.

3) If the observations belonging to the different groups are uncorrelated, we obtain

$m_{r_1,r_2}^{(i_1,i_2\div j,f)}=m_{r_1}^{(i_1,j)}\,m_{r_2}^{(i_2,f)}$

$\mu_{r_1,r_2}^{(i_1,i_2\div j,f)}=\mu_{r_1}^{(i_1,j)}\,\mu_{r_2}^{(i_2,f)}$ and so on and on the other hand

$\underline{M}_{r_1,r_2}^{(i_1,i_2)}=\underline{M}_{r_1}^{(i_1)}\,\underline{M}_{r_2}^{(i_2)}$

$M_{r_1,r_2}^{(i_1,i_2)}=M_{r_1}^{(i_1)}\,M_{r_2}^{(i_2)}$ and so on.

But in the general case (Cf. above chapter I, § I)

$M_{[r_1,r_2;S]}\neq M_{[r_1;S]}\;M_{[r_2;S]}$

$\underline{M}_{[r_1,r_2;S]}\neq\underline{M}_{[r_1;S]}\;\underline{M}_{[r_2;S]}$

§ II

1) Let us consider the quantities Z_i' as the variables which have to be dealt with and put:

$$E\left[Z_{(S)}\right]^r=M_{r,(S)}$$

$$E\left[Z_{(S)}-M_{1,(S)}\right]^r=\underline{M}_{r,(S)}\,.$$

$$\frac{1}{S}\sum_{i=1}^{S}\left[Z'_i - Z_{(S)}\right]^r = \underline{N}'_{[r;S]}$$

$$\boldsymbol{E}\, \underline{N}'_{[r;S]} = \underline{N}_{[r;S]}$$

We can use for the calculation of $M_{r;(S)}$, $\underline{M}_{r;(S)}$, $\underline{N}_{[r;S]}$ the formulae of the chapter I and of the chapter III, if we substitute therein for the quantities m the corresponding quantities M and for the quantities μ the corresponding quantities $\underline{M}$. We obtain thus from chapter I (5):

$$M_{1,(S)} = \boldsymbol{E}\, Z_{(S)} = M_{[1;S]}$$

$$M_{2,(S)} = M_{[1,1;S]} + \frac{1}{S}\left[M_{[2;S]} - M_{[1,1;S]}\right]$$

(1) $$M_{3,(S)} = M_{[1,1,1;S]} + \frac{3}{S}\left[M_{[2,1;S]} - M_{[1,1,1;S]}\right] + \frac{1}{S^2}\left[M_{[3;S]} - 3\,M_{[2,1;S]} + 2\,M_{[1,1,1;S]}\right]$$

$$M_{4,(S)} = M_{[1,1,1,1;S]} + \frac{6}{S}\left[M_{[2,1,1;S]} - M_{[1,1,1,1;S]}\right] + \frac{1}{S^2}\left[3\,M_{[2,2;S]} + 4\,M_{[3,1;S]} - 18\,M_{[2,1,1;S]} + 11\,M_{[1,1,1,1;S]}\right] + \frac{1}{S^3}\left[M_{[4;S]} - 3\,M_{[2,2;S]} - 4\,M_{[3,1;S]} + 12\,M_{[2,1,1;S]} - 6\,M_{[1,1,1,1;S]}\right]$$

We obtain similar formulae for $\underline{M}_{r,(S)}$, if we substitute in (1) for $M_{[r;S]}$ the corresponding quantities $\underline{M}_{[r;S]}$, for $M_{[r_1,r_2;S]}$ the corresponding quantities $\underline{M}_{[r_1,r_2;S]}$ and so on.

From chapter III (1), (2) and (8) we obtain similarly:

$$\underline{N}_{[2;S]} = \boldsymbol{E}\,\frac{1}{S}\sum_{i=1}^{S}\left[Z_i - Z_{(S)}\right]^2 = \frac{1}{S}\sum_{i=1}^{S}\left[M_1^{(i)} - M_{[1,S]}\right]^2 + \frac{S-1}{S}\left[\underline{M}_{[2;S]} - \underline{M}_{[1,1,S]}\right]$$

$$\underline{N}_{[3;S]}=\frac{1}{S}\sum_{i=1}^{S}\left[M_1^{(i)}-\underline{M}_{[1;S]}\right]^3+$$

$$+\frac{(S-1)\,(S-2)}{S^2}\left[\underline{M}_{[3;S]}-3\,\underline{M}_{[2,1;S]}+2\,\underline{M}_{[1,1,1;S]}\right]+$$

$$(2)\quad +\frac{3(S-2)}{S^2}\sum_{i=1}^{S}\underline{M}_2^{(i)}\left[M_1^{(i)}-\underline{M}_{[1;S]}\right]-\frac{6}{S^2}\sum_{i_1=1}^{S}\sum_{i_2\neq i_1}\underline{M}_{1,1}^{(i_1,i_2)}\left[M_1^{(i)}-\underline{M}_{[1;S]}\right]$$

$$\underline{N}_{[4;S]}=\frac{1}{S}\sum_{i=1}^{S}\left[M_1^{(i)}-\underline{M}_{[1;S]}\right]^4+$$

$$+\frac{(S-1)(S-2)(S-3)}{S^3}\left[\underline{M}_{[4;S]}-4\underline{M}_{[3,1;S]}+6\underline{M}_{[2,1,1;S]}-3\underline{M}_{[1,1,1,1;S]}\right]+$$

$$+\frac{(S-1)(2S-3)}{S^3}\left[\underline{M}_{[4;S]}-4\underline{M}_{[3,1;S]}+3\underline{M}_{[2,2;S]}\right]+$$

$$+\frac{6\,(S-2)}{S^2}\sum_{i=1}^{S}\underline{M}_2^{(i)}\left[M_1^{(i)}-\underline{M}_{[1;S]}\right]^2+\frac{6}{S}\underline{M}_{[2;S]}\left\{\frac{1}{S}\sum_{i=1}^{S}\left[M_1^{(i)}-\underline{M}_{[1;S]}\right]^2\right\}-$$

$$-\frac{12}{S^2}\sum_{i_1=1}^{S}\sum_{i_2\neq i_1}\underline{M}_{1,1}^{(i_1,i_2)}\left[M_1^{(i)}-\underline{M}_{[1;S]}\right]^2+$$

$$+\frac{6\,(S-1)}{S}\underline{M}_{[1,1;S]}\left\{\frac{1}{S}\sum_{i=1}^{S}\left[M_1^{(i)}-\underline{M}_{[1;S]}\right]^2\right\}+$$

$$+\frac{4\,(S^2-3\,S+3)}{S^3}\sum_{i=1}^{S}\underline{M}_3^{(i)}\left[M_1^{(i)}-\underline{M}_{[1;S]}\right]-$$

$$-\frac{12\,(S-2)}{S^3}\sum_{i_1=1}^{S}\sum_{i_2\neq i_1}\underline{M}_{2,1}^{(i_1,i_2)}\left[M_1^{(i_1)}-\underline{M}_{[1;S]}\right]+$$

$$+\frac{12}{S^3}\sum_{i_1=1}^{S}\sum_{i_2\neq i_1}\underline{M}_{1,2}^{(i_1,i_2)}\left[M_1^{(i_1)}-\underline{M}_{[1;S]}\right]+$$

$$+\frac{12}{S^3}\sum_{i_1=1}^{S}\sum_{i_2\neq i_1}\sum_{i_3\neq i_2\neq i_1}\underline{M}_{(1,1,1}^{i_1,i_2,i_3)}\left[M_1^{(i_1)}-\underline{M}_{[1;S]}\right]$$

$$\sigma^2_{\underline{N}'_{[2;S]}}=E\left\{\frac{1}{S}\sum_{i=1}^{S}\left[Z'_i-Z_{(s)}\right]^2-\left(\frac{1}{S}\sum_{i=1}^{S}\left[M_1^{(i)}-\underline{M}_{[1;S]}\right]^2+\right.\right.$$

$$\left.\left.+\frac{S-1}{S}\left[\underline{M}_{[2;S]}-\underline{M}_{[1,1;S]}\right]\right)\right\}^2=\frac{(S-1)^2}{S^3}\left[\underline{M}_{[4;S]}-4\underline{M}_{[3,1;S]}+3\underline{M}_{[2,2,S]}\right]+$$

$$(3)\quad +\frac{(S-1)\,(S-2)\,(S-3)}{S^3}\left[\underline{M}_{[2,2;S]}-2\underline{M}_{[2,1,1;S]}+\underline{M}_{[1,1,1,1;S]}\right]-$$

$$-\frac{(S-1)^2}{S^2}\left[\underline{M}_{[2;S]}-\underline{M}_{[1,1;S]}\right]^2+\frac{4}{S^2}\sum_{i=1}^{S}\underline{M}_2^{(i)}\left[M_1^{(i)}-M_{[1;S]}\right]^2+$$

$$+\frac{4}{S^2}\sum_{i_1=1}^{S}\sum_{i_2\neq i_1}\underline{M}_{1,1}^{(i_1,i_2)}\left[M_1^{(i_1)}-M_{[1;S]}\right]\left[M_1^{(i_2)}-M_{[1;S]}\right]+$$

$$+\frac{4(S-1)}{S^3}\sum_{i=1}^{S}\underline{M}_3^{(i)}\left[M_1^{(i)}-M_{[1;S]}\right]+$$

$$+\frac{4\,(S-1)}{S^3}\sum_{i_1=1}^{S}\sum_{i_2\neq i_1}\underline{M}_{2,1}^{(i_1,i_2)}\left[M_1^{(i_2)}-M_{[1;S]}\right]-$$

$$-\frac{8}{S^3}\sum_{i_1=1}^{S}\sum_{i_2\neq i_1}\underline{M}_{2.1}^{(i_1,i_2)}\left[M_1^{(i_1)}-M_{[1;S]}\right]-$$

$$-\frac{4}{S^3}\sum_{i_1=1}^{S}\sum_{i_2\neq i_1}\sum_{i_3\neq i_2\neq i_1}\underline{M}_{1,1,1}^{(i_1,i_2,i_3)}\left[M_1^{(i_1)}-M_{[1;S]}\right]$$

2) If the numbers $n_1, n_2, \ldots, n_s$ are unequal, the unweighted average of the quantities Z'_i - i.e. $Z_{(S)}$ - does not coincide with their weighted average - i.e. $X_{(N)}=\frac{1}{N}\sum_{i=1}^{S} n_i Z'_i$ - and it may be preferable to calculate the deviations of the quantities Z'_i from $X_{(N)}$.

Noting that

$$\boldsymbol{E}\,X^2_{(N)}=\frac{1}{N^2}\left\{\sum_{i=1}^{S}(n_i)^2 M_2^{(i)}+\sum_{i_1=1}^{S}\sum_{i_2\neq i_1} n_{i_1} n_{i_2} M_{1,1}^{(i_1,i_2)}\right\}$$

we find:

$$\boldsymbol{E}\,\frac{1}{N}\sum_{i=1}^{S} n_i\left[Z'_i-X_{(N)}\right]^2=\boldsymbol{E}\,\frac{1}{N}\sum_{i=1}^{S} n_i\,(Z'_i)^2-\boldsymbol{E}\,X^2_{(N)}=$$

$$(4)\qquad =\frac{1}{N^2}\sum_{i=1}^{S} n_i\,(N-n_i)\,M_2^{(i)}-\frac{1}{N^2}\sum_{i_1=1}^{S}\sum_{i_2\neq i_1} n_{i_1} n_{i_2} M_{1,1}^{(i_1,i_2)}$$

$$\sigma^2_{\frac{1}{N}\sum_{i=1}^{S} n_i\left[Z'_i-X_{(N)}\right]^2}=\boldsymbol{E}\left\{\frac{1}{N}\sum_{i=1}^{S} n_i\left[Z'_i-X_{(N)}\right]^2-\left[\frac{1}{N^2}\sum_{i=1}^{S} n_i(N-n_i)\,M_2^{(i)}-\right.\right.$$

$$\left.\left.-\frac{1}{N^2}\sum_{i_1=1}^{S}\sum_{i_2\neq i_1} n_{i_1} n_{i_2} M_{1,1}^{(i_1,i_2)}\right]\right\}^2=$$

$$=\frac{1}{N^4}\sum_{i=1}^{S} n_i^2(N-n_i)^2\left[M_4^{(i)}-(M_2^{(i)})^2\right]+$$

$$+\frac{1}{N^4}\sum_{i_1=1}^{S}\sum_{i_2\neq i_1} n_{i_1} n_{i_2}(N-n_{i_1})(N-n_{i_2})\left[M_{2,2}^{(i_1,i_2)}-M_2^{(i_1)} M_2^{(i_2)}\right]+$$

$$(5) \quad + \frac{2}{N^4} \sum_{i_1=1}^{S} \sum_{i_2 \neq i_1} (n_{i_1})^2 (n_{i_2})^2 \left[M_{2,2}^{(i_1,i_2)} - \left(M_{1,1}^{(i_1,i_2)}\right)^2 \right] -$$

$$- \frac{4}{N^4} \sum_{i_1=1}^{S} \sum_{i_2 \neq i_1} (n_{i_1})^2 n_{i_2} (N - n_{i_1}) \left[M_{3,1}^{(i_1,i_2)} - M_2^{(i_1)} M_{1,1}^{(i_1,i_2)} \right] -$$

$$- \frac{2}{N^4} \sum_{i_1=1}^{S} \sum_{i_2 \neq i_1} \sum_{i_3 \neq i_2 \neq i_1} n_{i_1} n_{i_2} n_{i_3} (N - n_{i_1}) \left[M_{2,1,1}^{(i_1,i_2,i_3)} - M_2^{(i_1)} M_{1,1}^{(i_1,i_2)} \right] +$$

$$+ \frac{4}{N^4} \sum_{i_1=1}^{S} \sum_{i_2 \neq i_1} \sum_{i_3 \neq i_2 \neq i_1} (n_{i_1})^2 n_{i_2} n_{i_3} \left[M_{2,1,1}^{(i_1,i_2,i_3)} - M_{1,1}^{(i_1 i_2)} M_{1,1}^{(i_1,i_3)} \right] +$$

$$+ \frac{1}{N^4} \sum_{i_1=1}^{S} \sum_{i_2 \neq i_1} \sum_{i_3 \neq i_2 \neq i_1} \sum_{i_4 \neq i_3 \neq i_2 \neq i_1} n_{i_1} n_{i_2} n_{i_3} n_{i_4} \left[M_{1,1,1,1}^{(i_1,i_2,i_3,i_4)} - M_{1,1}^{(i_1,i_2)} M_{1,1}^{(i_3,i_4)} \right]$$

§ III

The general formulae of § II are for themselves of no special interest; their use is an auxiliary one, -to make the analysis of the specialized problems more easy, as they permit to spare in each single case the deductions which result in the general case in the unwieldy formulae of § II. The single problems can be, of course, more or less specialized. I shall analyze below (chapter V and chapter VI) some problems which are of actual statistical interest. The following formulae shall further facilitate hereby the calculations.

Suppose that all quantities m_r, all quantities μ_r, all quantities μ_{r_1,r_2} and so on are equal. Substituting the corresponding values of $M_r^{(i)}$, $M_{[r;S]}$, $M_{r_1,r_2}^{(i_1,i_2)}$ $M_{[r_1,r_2;S]}$ and so on (see above § I,2), we obtain:

$$\boldsymbol{E} z_{(S)} = m_1$$

$$\boldsymbol{E}\left[z_{(S)} - m_1 \right]^2 = \mu_{1,1} + \frac{1}{S} \left[\frac{1}{S} \sum_{i=1}^{S} \frac{1}{n_i} \right] \left[\mu_2 - \mu_{1,1} \right]$$

$$(6) \quad \boldsymbol{E}\left[z_{(S)} - m_1 \right]^3 = \mu_{1,1,1} + \frac{3}{S} \left[\frac{1}{S} \sum_{i=1}^{S} \frac{1}{n_i} \right] \left[\mu_{2,1} - \mu_{1,1,1} \right] +$$

$$+ \frac{1}{S^2} \left[\frac{1}{S} \sum_{i=1}^{S} \frac{1}{(n_i)^2} \right] \left[\mu_3 - 3\,\mu_{2,1} + 2\,\mu_{1,1,1} \right]$$

$$\boldsymbol{E}\left[z_{(S)} - m_1 \right]^4 = \mu_{1,1,1,1} + \frac{6}{S} \left[\frac{1}{S} \sum_{i=1}^{S} \frac{1}{n_i} \right] \left[\mu_{2,1,1} - \mu_{1,1,1,1} \right] +$$

$$+\frac{3}{S}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{n_i}\right]^2\left[\mu_{2,2}-2\,\mu_{2,1,1}+\mu_{1,1,1,1}\right]+$$

$$+\frac{1}{S^2}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{(n_i)^2}\right]\left[4\,\mu_{3,1}-3\,\mu_{2,2}-6\,\mu_{2,1,1}+5\,\mu_{1,1,1,1}\right]+$$

$$+\frac{3}{S^2}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{n_i}\right]^2\left[\mu_{2,2}-2\,\mu_{2,1,1}+\mu_{1,1,1,1}\right]+$$

$$+\frac{6}{S^3}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{(n_i)^2}\right]\left[\mu_{2,2}-2\,\mu_{2,1,1}+\mu_{1,1,1,1}\right]+$$

$$+\frac{1}{S^3}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{(n_i)^3}\right]\left[\mu_4-4\,\mu_{3,1}-3\,\mu_{2,2}+12\,\mu_{2,1,1}-6\,\mu_{1,1,1,1}\right]$$

$$\boldsymbol{E}\,\frac{1}{S}\sum_{i=1}^{S}\left[z'_i-z_{(S)}\right]^2=\frac{S-1}{S}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{n_i}\right]\left[\mu_2-\mu_{1,1}\right]$$

$$(7)\quad \boldsymbol{E}\,\frac{1}{S}\sum_{i=1}^{S}\left[z'_i-z_{(S)}\right]^3=$$

$$=\frac{(S-1)\,(S-2)}{S^2}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{(n_i)^2}\right]\left[\mu_3-3\,\mu_{2,1}+2\,\mu_{1,1,1}\right]$$

$$\boldsymbol{E}\,\frac{1}{S}\sum_{i=1}^{S}\left[z'_i-z_{(S)}\right]^4=$$

$$=\frac{(S-1)\,(S^2-3\,S+3)}{S^3}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{(n_i)^3}\right]\left[\mu_4-4\mu_{3,1}-3\mu_{2,2}+12\mu_{2,1,1}-\right.$$

$$\left.-6\mu_{1,1,1,1}\right]+\left\{\frac{3(S-2)^2}{S^2}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{(n_i)^2}\right]+\right.$$

$$\left.+\frac{3\,(2S-3)}{S^2}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{n_i}\right]^2\right\}\left[\mu_{2,2}-2\mu_{2,1,1}+\mu_{1,1,1,1}\right]$$

$$\boldsymbol{E}\left\{\frac{1}{S}\sum_{i=1}^{S}\left[z'_i-z_{(S)}\right]^2-\frac{S-1}{S}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{n_i}\right]\left[\mu_2-\mu_{1,1}\right]\right\}^2=$$

$$(8)\quad =\frac{(S-1)^2}{S^3}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{(n_i)^3}\right]\left[\mu_4-4\mu_{3,1}-3\mu_{2,2}+12\mu_{2,1,1}-6\mu_{1,1,1,1}\right]+$$

$$+\left\{\frac{S^2-2\,S+3}{S^2}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{n_i}\right]^2+\right.$$

$$\left.+\frac{2(S-2)}{S^2}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{(n_i)^2}\right]\right\}\left[\mu_{2,2}-2\mu_{2,1,1}+\mu_{1,1,1,1}\right]-$$

$$- \frac{(S-1)^2}{S^2} \left[\frac{1}{S} \sum_{i=1}^{S} \frac{1}{n_i} \right]^2 \left[\mu_2 - \mu_{1,1} \right]^2$$

$$(9) \qquad \mathbf{E} \frac{1}{N} \sum_{i=1}^{S} n_i \left[Z_i' - X_{(N)} \right]^2 = \frac{S-1}{N} \left[\mu_2 - \mu_{1,1} \right]$$

$$\mathbf{E} \left\{ \frac{1}{N} \sum_{i=1}^{S} n_i \left[Z_i' - X_{(N)} \right]^2 - \frac{S-1}{N} \left[\mu_2 - \mu_{1,1} \right] \right\}^2 =$$

$$= \frac{1}{N^4} \left[\sum_{i=1}^{S} \frac{(N-n_i)^2}{n_i} \right] \left[\mu_4 - 4\mu_{3,1} - 3\mu_{2,2} + 12\mu_{2,1,1} - 6\mu_{1,1,1,1} \right] +$$

$$(10) \quad + \frac{(S-1)(S+1)}{N^2} \left[\mu_{2,2} - 2\mu_{2,1,1} + \mu_{1,1,1,1} \right] - \frac{(S-1)^2}{N^2} \left[\mu_2 - \mu_{1,1} \right]^2$$

CHAPTER V

§ I

1) From a closed urn which contains a_1 tickets marked with the number ξ_1, a_2 tickets marked with the number ξ_2, and so on, a_k tickets marked with the number ξ_k, α tickets are drawn successively whereby the tickets which happen to be drawn are not replaced in the urn. Setting $a_1+a_2+....+a_k=A$, put:

$$\sum_{j=1}^{K} \frac{a_j}{A} \xi_j^r = m_r$$

$$\sum_{h=0}^{r} (-1)^h C_r^h m_1^h m_{r-h} = \mu_r$$

Denoting by X_i the number marked on the ticket which happens to be drawn at the i-*th* drawing, we find without difficulty that

$$\mathbf{E}\, X_1^r = m_r$$

$$\mathbf{E} \left[X_1 - m_1 \right]^r = \mu_r.$$

In the chapter III of my paper on « *Uber den mittleren Fehler des Durchschnittes von gegenseitig nicht unabhängigen Grössen* » (*) I have demonstrated that

(*) Cf. A. Tschuprow, *Zur Theorie der Stabilität statistischer Reihen*, S. 216-219 « Skandinavisk Aktuarietidskrift » 1918.

$$\boldsymbol{E}\, X_i = \boldsymbol{E}\, X_1 = m_1$$

$$\boldsymbol{E}\, X_i^2 = \boldsymbol{E}\, X_1^2 = m_2$$

$$\boldsymbol{E}\, X_i\, X_j = \boldsymbol{E}\, X_1\, X_2 = \frac{A}{A-1}\, m_1^2 - \frac{1}{A-1}\, m_2$$

$$\boldsymbol{E}\, X_i\, X_j - \boldsymbol{E}\, X_i\, \boldsymbol{E}\, X_j = -\frac{1}{A-1}\, \mu_2 .$$

Putting

$$\boldsymbol{E}\, X_i^r = m_r^{(i)}, \boldsymbol{E}\, X_i^{r_1}\, X_j^{r_2} = m_{r_1,r_2}^{(i,j)},\ \boldsymbol{E}\, X_i^{r_1} X_j^{r_2} X_h^{r_3} = m_{r_1,r_2,r_3}^{(i,j,h)} \quad \text{and so on}$$

$$\boldsymbol{E}\left[X_i - m_1\right]^r = \mu_r^{(i)},\ \boldsymbol{E}\left[X_i - m_1\right]^{r_1}\left[X_j - m_1\right]^{r_2} = \mu_{r_1,r_2}^{(i,j)},$$

$$\boldsymbol{E}\left[X_i - m_1\right]^{r_1}\left[X_j - m_1\right]^{r_2}\left[X_h - m_1\right]^{r_3} = \mu_{r_1,r_2,r_3}^{(i,j,h)},$$

we have consequently:

$$m_{1,1}^{(i,j)} = m_{1,1}^{(1,2)} = \frac{A}{A-1}\, m_1^2 - \frac{1}{A-1}\, m_2$$

$$\mu_{1,1}^{(i,j)} = \mu_{1,1}^{(1,2)} = -\frac{1}{A-1}\, \mu_2$$

Similarly we find:

$$m_{1,1,1}^{(1,2,3)} = \boldsymbol{E}\, X_1 X_2 X_3 = \sum_{j=1}^{K} \frac{a_j\,(a_j-1)}{A(A-1)}\, \xi_j^2 \left[\frac{a_j-2}{A-2}\, \xi_j + \sum_{h \neq j} \frac{a_h}{A-2}\, \xi_h\right] +$$

$$+ \sum_{j=1}^{K} \sum_{i \neq j} \frac{a_j\, a_i}{A(A-1)}\, \xi_j\, \xi_i \left[\frac{a_j-1}{A-2}\, \xi_j + \frac{a_i-1}{A-2}\, \xi_i + \sum_{h \neq j \neq i} \frac{a_h}{A-2}\, \xi_h\right] =$$

$$= \frac{1}{A(A-1)(A-2)} \left\{ \left[\sum_{j=1}^{K} a_j\, \xi_j\right]^3 - 3\left[\sum_{j=1}^{K} a_j\, \xi_j\right]\left[\sum_{j=1}^{K} a_j\, \xi_j^2\right] + \right.$$

$$\left. + 2 \sum_{j=1}^{K} a_j\, \xi_j^3 \right\} = \frac{1}{(A-1)\,(A-2)} \left\{ A^2\, m_1^3 - 3\, A\, m_1\, m_2 + 2\, m_3 \right\}$$

$$m_{2,1}^{(1,2)} = m_{1,2}^{(1,2)} = \frac{1}{A-1} \left\{ A\, m_1\, m_2 - m_3 \right\}$$

$$m_{1,1,1,1}^{(1,2,3,4)} = \frac{1}{(A-1)(A-2)(A-3)} \cdot$$

$$\cdot \left\{ A^3\, m_1^4 - 6\, A^2\, m_1^2\, m_2 + 8\, A\, m_1\, m_3 + 3\, A\, m_2^2 - 6\, m_4 \right\}$$

$$m^{(1,2,3)}_{2,1,1} = m^{(1,2,3)}_{1,2,1} = m^{(1,2,3)}_{1,1,2} = \frac{1}{(A-1)\,(A-2)} \cdot$$

$$\cdot \left\{ A^2 m_1^{\,2} m_2 - 2\,A m_1 m_3 - A m_2^{\,2} + 2\,m_4 \right\}$$

$$m^{(1,2)}_{2,2} = \frac{1}{A-1} \left\{ A m_2^{\,2} - m_4 \right\}$$

$$m^{(1,2)}_{3,1} = m^{(1,2)}_{1,3} = \frac{1}{A-1} \left\{ A m_1 m_3 - m_4 \right\}$$

and hence

$$\mu^{(1,2,3)}_{1,1,1} = \frac{2}{(A-1)(A-2)}\,\mu_3$$

$$\mu^{(1,2)}_{2,1} = \mu^{(1,2)}_{1,2} = -\frac{1}{A-1}\,\mu_3$$

$$\mu^{(1,2,3,4)}_{1,1,1,1} = -\frac{6}{(A-1)\,(A-2)\,(A-3)}\,\mu_4 + \frac{3A}{(A-1)\,(A-2)\,(A-3)}\,\mu_2^2$$

$$\mu^{(1,2,3)}_{2,1,1} = \mu^{(1,2,3)}_{1,2,1} = \mu^{(1,2,3)}_{1,1,2} = \frac{2}{(A-1)\,(A-2)}\,\mu_4 - \frac{A}{(A-1)\,(A-2)}\,\mu_2$$

$$\mu^{(1,2)}_{2,2} = -\frac{1}{A-1}\,\mu_4 + \frac{A}{A-1}\,\mu_2^2$$

$$\mu^{(1,2)}_{3,1} = \mu^{(1,2)}_{1,3} = -\frac{1}{A-1}\,\mu_4$$

Let us denote by $(m_r^{(2)})_j$ the value which $m_r^{(2)}$ assumes, if a ticket marked ξ_j happens to be drawn at the first drawing. Noting that

$$\left(m_r^{(2)}\right)_j = \frac{a_j-1}{A-1}\,\xi_j^r + \sum_{h\neq j} \frac{a_h}{A-1}\,\xi_h^r = \frac{A}{A-1}\,m_r - \frac{1}{A-1}\,\xi_j^r\,;$$

we find:

$$m_r^{(2)} = \sum_{j=1}^{K} \frac{a_j}{A} \left(m_r^{(2)}\right)_j = \frac{A}{A-1}\,m_r - \frac{1}{A-1}\,m_r = m_r$$

$$m^{(1,2)}_{r_1,r_2} = \sum_{j=1}^{K} \frac{a_j}{A}\,\xi_j^{r_1} \left(m^{(2)}_{r_2}\right)_j = \frac{A}{A-1}\,m_{r_1} m_{r_2} - \frac{1}{A-1}\,m_{r_1+r_2}$$

Similarly we find:

$$\left(m^{(2,3)}_{r_1,r_2}\right)_j = \frac{A-1}{A-2}\left(m_{r_1}\right)_j\left(m_{r_2}\right)_j - \frac{1}{A-2}\left(m_{r_1+r_2}\right)_j =$$

$$= \frac{1}{(A-1)(A-2)}\left\{A^2 m_{r_1} m_{r_2} - A\left[m_{r_1+r_2} + m_{r_1}\xi_j^{r_2} + m_{r_2}\xi_j^{r_1}\right] + 2\xi_j^{r_1+r_2}\right\}$$

and hence

$$m^{(2,3)}_{r_1,r_2} = m^{(1,2)}_{r_1,r_2}$$

$$m^{(1,2,3)}_{r_1;r_2,r_3} = \frac{1}{(A-1)(A-2)}\cdot$$

$$\cdot\left\{A^2 m_{r_1} m_{r_2} m_{r_3} - A\left[m_{r_1} m_{r_2+r_3} + m_{r_2} m_{r_1+r_3} + m_{r_3} m_{r_1+r_2}\right] + 2 m_{r_1+r_2+r_3}\right\}$$

It is easy to demonstrate in the same manner that

$$m^{(i)}_r = m_r\,,\ m^{(i,j)}_{r_1,r_2} = m^{(1,2)}_{r_1,r_2},\ m^{(i,j,h)}_{r_1,r_2,r_3,} = m^{(1,2,3)}_{r_1,r_2,r_3} \text{ and so on}$$

$$\mu^{(i)}_r = \mu_r,\ \mu^{(i,j)}_{r_1,r_2} = \mu^{(1,2)}_{r_1,r_2} \text{ and so on}$$

On the other hand it is easy to see that

$$m_{r_1,r_2,r_3,r_4,} = \frac{1}{(A-1)(A-2)(A-3)}\Big\{A^3 m_{r_1} m_{r_2} m_{r_3} m_{r_4} -$$

$$- A^2\Big[m_{r_1} m_{r_2} m_{r_3+r_4} + m_{r_1} m_{r_3} m_{r_2+r_4} + m_{r_1} m_{r_4} m_{r_2+r_3} + m_{r_2} m_{r_3} m_{r_1+r_4} +$$

$$+ m_{r_2} m_{r_4} m_{r_1+r_3} + m_{r_3} m_{r_4} m_{r_1+r_2}\Big] + A\Big[m_{r_1+r_2} m_{r_3+r_4} + m_{r_1+r_3} m_{r_2+r_4} +$$

$$+ m_{r_1+r_4} m_{r_2+r_3} + 2\Big(m_{r_1} m_{r_2+r_3+r_4} + m_{r_2} m_{r_1+r_3+r_4} + m_{r_3} m_{r_1+r_2+r_4} +$$

$$+ m_{r_4} m_{r_1+r_2+r_3}\Big)\Big] - 6\, m_{r_1+r_2+r_3+r_4}\Big\}$$

and that in general

$$m_{r_1,r_2,\dots,r_k} = \frac{1}{A^{[-k]}}\Big\{A^k m_{r_1} m_{r_2} \cdots m_{r_k} -$$

$$- A^{k-1}\Big[m_{r_1} m_{r_2} \cdots m_{r_{k-2}} m_{r_{k-1}+r_k} + \cdots\Big] +$$

$$+ A^{k-2}\left[\left(m_{r_1} m_{r_2} \cdots m_{r_{k-4}} m_{r_{k-3}+r_{k-2}} m_{r_{k-1}+r_k} + \cdots\right) + \right.$$

$$\left. + 2\left(m_{r_1} m_{r_2} \cdots m_{r_{k-3}} m_{r_{k-2}+r_{k-1}+r_k} + \cdots\right)\right] -$$

$$- \cdots + \left(-1\right)^{k-1} A \,.\, 1 \,.\, 2 \ldots (k-1)\, m_{r_1+r_2+\cdots+r_k} \Big\}$$

Substituting for the quantities m the corresponding quantities μ and noting that $\mu_1 = 0$, we obtain the formulae for

$\mu_{r_1,r_2}, \mu_{r_1,r_2,r_3}$ and so on. We have, for example,

$$\mu_{r,1} = -\frac{1}{A-1}\,\mu_{r+1}$$

$$\mu^{\{2r\}}_{1,1,\cdots,1} = \frac{(-1)^r}{A^{[-2r]}}\Big\{A^r 1 \,.\, 3 \,.\, 5 \ldots (2r-1)\,\mu_2^r + \cdots\Big\}$$

$$\mu^{\{2r+1\}}_{1,1,\cdots,1} = \frac{(-1)^{r+1}}{A^{[-(2r+1)]}}\Big\{A^r \frac{2r}{3}\, 1 \,.\, 3 \,.\, 5 \ldots (2r+1)\,\mu_2^{r-1}\mu_3 + \cdots\Big\}$$

2) Substituting the values of $\mu_2^{(i)}$, $\mu_{1,1}^{(i,j)}$ an so on found above in (5) and (9) of the chapter I and in (3), (7) and (9) of the chapter III, we find after due transformations: (*)

$$m_{2;(\alpha)} = m_1^2 + \frac{A-\alpha}{(A-1)\,\alpha}\,\mu_2$$

$$(1) \qquad m_{3(\alpha)} = m_1^3 + 3\,\frac{A-\alpha}{(A-1)\,\alpha}\, m_1\,\mu_2 + \frac{1}{\alpha^2}\left[1-3\frac{\alpha-1}{A-1} + 2\frac{(\alpha-1)\,(\alpha-2)}{(A-1)(A-2)}\right]\mu_3$$

$$m_{4,(\alpha)} = m_1^4 + 6\frac{A-\alpha}{(A-1)\,\alpha} m_1^2\mu_2 + 4\frac{1}{\alpha^2}\left[1-3\frac{\alpha-1}{A-1} + 2\frac{(\alpha-1)(\alpha-2)}{(A-1)(A-2)}\right] m_1\,\mu_3 +$$

$$+ 3\,\frac{A\,(\alpha-1)}{(A-1)\alpha^3}\left[1-2\,\frac{\alpha-2}{A-2} + \frac{(\alpha-2)\,(\alpha-3)}{(A-2)(A-3)}\right]\mu_2^2 +$$

$$+ \frac{1}{\alpha^3}\left[1-7\,\frac{\alpha-1}{A-1} + 12\,\frac{(\alpha-1)\,(\alpha-2)}{(A-1)(A-2)} - 6\,\frac{(\alpha-1)(\alpha-2)(\alpha-3)}{(A-1)(A-2)\,(A-3)}\right]\mu_4$$

(*) The general formulae being excessively unwieldy, I do limit myself in the general case to the calculation of $m_{r,(\alpha)}$ for $r = 2,3,\ 4$. I give below (§I,3) the general formula for $\mu_{r,(\alpha)}$ in the case, when A and α are great numbers of the same order and in the chapter VI (11) and (12) the general formulae for $m_{r,(\alpha)}$ an $\mu_{r,(\alpha)}$ in the special case of statistical frebuencies.

$$\mu_{2,(\alpha)} = \frac{A-\alpha}{A-1}\frac{1}{\alpha}\mu_2 = \left[1 - \frac{\alpha-1}{A-1}\right]\frac{1}{\alpha}\mu_2$$

$$(2) \qquad \mu_{3,(\alpha)} = \frac{1}{\alpha^2}\left[1-3\frac{\alpha-1}{A-1} + 2\frac{(\alpha-1)(\alpha-2)}{(A-1)(A-2)}\right]\mu_3 = \frac{(A-\alpha)(A-2\alpha)}{(A-1)(A-2)}\frac{1}{\alpha^2}\mu_3$$

$$\mu_{4,(\alpha)} = \frac{1}{\alpha^3}\left\{\left[1-7\frac{\alpha-1}{A-1} + 12\frac{(\alpha-1)(\alpha-2)}{(A-1)(A-2)} - 6\frac{(\alpha-1)(\alpha-2)(\alpha-3)}{(A-1)(A-2)(A-3)}\right]\mu_4 + \right.$$

$$\left. + 3\frac{A(\alpha-1)}{A-1}\left[1-2\frac{\alpha-2}{A-2} + \frac{(\alpha-2)(\alpha-3)}{(A-2)(A-3)}\right]\mu_2^2\right\} =$$

$$= \frac{A-\alpha}{\alpha^3}\left\{\frac{(A-2\alpha)(A-3\alpha)-A(\alpha-1)}{(A-1)(A-2)(A-3)}\mu_4 + 3\frac{A(\alpha-1)(A-\alpha-1)}{(A-1)(A-2)(A-3)}\mu_2^2\right\} \; (*)$$

$$v_{[2;\alpha]} = E\frac{1}{\alpha}\sum_{i=1}^{\alpha}\left[X_i - X_{(\alpha)}\right]^2 = \frac{\alpha-1}{\alpha}\frac{A}{A-1}\mu_2$$

$$v_{[3;\alpha]} = \frac{(\alpha-1)(\alpha-2)}{\alpha^2}\frac{A^2}{(A-1)(A-2)}\mu_3$$

$$(3) \qquad v_{[4;\alpha]} = \frac{(\alpha-1)(\alpha-2)(\alpha-3)}{\alpha^3}\frac{A}{(A-1)(A-2)(A-3)}$$

$$\left[(A^2-2A+3)\mu_4 - 3(2A-3)\mu_2^2\right] + \frac{(\alpha-1)(2\alpha-3)}{\alpha^3}\frac{A}{A-1}\left[\mu_4 + 3\mu_2^2\right]$$

$$E\left[v'_{[2;\alpha]}\right]^2 = E\left\{\frac{1}{\alpha}\sum_{i=1}^{\alpha}\left[X_i - X_{(\alpha)}\right]^2\right\}^2 = \frac{(\alpha-1)A(A-\alpha)\left[\alpha A - A - \alpha - 1\right]}{\alpha^3(A-1)(A-2)(A-3)}\mu_4 +$$

$$(4) \qquad + \frac{(\alpha-1)A\left[\alpha(\alpha+1)A^2 - 3(A-1)(\alpha^2 + \alpha A - A + \alpha)\right]}{\alpha^3(A-1)(A-2)(A-3)}\mu_2^2$$

$$\sigma^2_{v'_{[2;\alpha]}} = E\left[\frac{1}{\alpha}\sum_{i=1}^{\alpha}\left[X_i - X_{(\alpha)}\right]^2 - \frac{\alpha-1}{\alpha}\frac{A}{A-1}\mu_2\right]^2 =$$

$$(5) \qquad = \frac{(\alpha-1)A(A-\alpha)}{\alpha^3(A-1)(A-2)(A-3)}\left\{\left[\alpha A - A - \alpha - 1\right]\mu_4 - \right.$$

(*) Cf. L. Isserlis, *On the conditions under which the* «Probable errors» *of frequency distributions have real significance*, p. 30,31 «Proc. R. Soc.» A, vol. 92, London, 1916, (note that to my $\mu_{2,(\alpha)}$, $\mu_3(\alpha)$ and $\mu_4(\alpha)$ correspond Isserlis M_2', M_3' and M_4' for $\nu = 1$); L. Isserlis, *On the value of a mean, as calculated from a sample* «J. R. Stat. Soc.», vol. 81, London, 1918; G. Mortara, *Elementi di statistica*, p. 356 Roma, 1917; A. Tschuprow, *Zur Theorie der Stabilität statistischer Reihen*, p. 219 «Skandinavisk Aktuarietidskrift, 1918».

$$-\frac{1}{A-1}\left[\alpha A^2-3A^2+6A-3\alpha-3\right]\mu_2^2\Big\}=$$

$$=\frac{(\alpha-1)A(A-\alpha)}{\alpha^3(A-1)(A-2)(A-3)}\Big\{\left[\alpha A-A-\alpha-1\right]\left[\mu_4-3\mu_2^2\right]+$$

$$+\frac{2}{A-1}\left[\alpha A^2-3(\alpha+1)(A-1)\right]\mu_2^2\Big\}$$

3) With increasing A and constant α (1)-(5) tend to assume the form of (11) and (26) of the chapter I and of (7) and (23) of the chapter III of my paper « *On the math. exp.* ». Part I (*) If A and α are great numbers of the same order, we have approximately:

$$m_{2,(\alpha)}=m_1^2+\left[\frac{1}{\alpha}-\frac{1}{A}\right]\mu_2$$

$$m_{3,(\alpha)}=m_1^3+3\left[\frac{1}{\alpha}-\frac{1}{A}\right]m_1\mu_2+\frac{1}{\alpha^2}\left[1-3\frac{\alpha}{A}+2\frac{\alpha^2}{A^2}\right]\mu_3$$

$$m_{4,(\alpha)}=m_1^4+6\left[\frac{1}{\alpha}-\frac{1}{A}\right]m_1^2\mu_2+4\frac{1}{\alpha^2}\left[1-3\frac{\alpha}{A}+2\frac{\alpha^2}{A^2}\right]m_1\mu_3+$$

$$+3\frac{1}{\alpha^2}\left[1-2\frac{\alpha}{A}+\frac{\alpha^2}{A^2}\right]\mu_2^2+\frac{1}{\alpha^3}\left[1-7\frac{\alpha}{A}+12\frac{\alpha^2}{A^2}-6\frac{\alpha^3}{A^3}\right]\mu_4$$

$$\mu_{2,(\alpha)}=\frac{A-\alpha}{A\alpha}\mu_2=\left[\frac{1}{\alpha}-\frac{1}{A}\right]\mu_2=\frac{1}{\alpha}\left(1-\frac{\alpha}{A}\right)\mu_2$$

$$\mu_{3,(\alpha)}=\frac{1}{\alpha^2}\left[1-3\frac{\alpha}{A}+2\frac{\alpha^2}{A^2}\right]\mu_3=\frac{1}{\alpha^2}\left[1-\frac{\alpha}{A}\right]\left[1-2\frac{\alpha}{A}\right]\mu_3=$$

$$=\left[\frac{1}{\alpha}-\frac{1}{A}\right]\left[\frac{1}{\alpha}-\frac{2}{A}\right]\mu_3$$

$$\mu_{4,(\alpha)}=\frac{1}{\alpha^3}\Big\{\left[1-7\frac{\alpha}{A}+12\frac{\alpha^2}{A^2}-6\frac{\alpha^3}{A^3}\right]\mu_4+$$

$$+3\alpha\left[1-2\frac{\alpha}{A}+\frac{\alpha^2}{A^2}\right]\mu_2^2\Big\}=$$

$$=\frac{1}{\alpha^3}\left[1-\frac{\alpha}{A}\right]\left[1-6\frac{\alpha}{A}\left(1-\frac{\alpha}{A}\right)\right]\mu_4+3\frac{1}{\alpha^2}\left[1-\frac{\alpha}{A}\right]^2\mu_2^2=$$

(*) « Biometrika » vol. XII pp. 151, 155, 186, 193.

$$=\left[\frac{1}{a}-\frac{1}{A}\right]\left\{\left[\left(\frac{1}{a}-\frac{2}{A}\right)\left(\frac{1}{a}-\frac{3}{A}\right)-\frac{1}{Aa}\right]\mu_4+3\left[\frac{1}{a}-\frac{1}{A}\right]\mu_2^2\right\}$$

$$\mu_{2r,(a)}=1.3.5\ldots.(2r-1)\frac{1}{a^r}\left(1-\frac{a}{A}\right)^r\mu_2^r+\ldots..$$

$$\mu_{2r+1,(a)}=1.3.5\ldots(2r+1)\frac{r}{3}\frac{1}{a^{r+1}}\left(1-\frac{a}{A}\right)^r\left(1-\frac{2a}{A}\right)\mu_2^{r-1}\mu_3+\ldots..$$

$$v_{[2;a]}=\mu_2$$

$$v_{[3;a]}=\mu_3$$

$$v_{[4;a]}=\left[1+2\left(\frac{1}{a}-\frac{1}{A}\right)\right]\mu_4+6\left[\frac{1}{a}-\frac{1}{A}\right]\mu_2^2=$$

$$=\mu_4+2\left[\frac{1}{a}-\frac{1}{A}\right]\left[\mu_4+3\mu_2^2\right].$$

$$\sigma^2_{v'[2;a]}=\frac{A-a}{A\,a}\left[\mu_4-\mu_2^2\right]=\left[\frac{1}{a}-\frac{1}{A}\right]\left[\mu_4-\mu_2^2\right]$$

Noting that

$$\frac{\mu_{3,(a)}^2}{\mu_{2,(a)}^3}=\frac{\left[\frac{1}{a}-\frac{2}{A}\right]^2}{\left[\frac{1}{a}-\frac{1}{A}\right]}\frac{\mu_3^2}{\mu_2^3}$$

$$\frac{\mu_{4,(a)}}{\mu_{2,(a)}^2}=3+\frac{\left[\frac{1}{a}-\frac{2}{A}\right]\left[\frac{1}{a}-\frac{3}{A}\right]-\frac{1}{Aa}}{\left[\frac{1}{a}-\frac{1}{A}\right]}\frac{\mu_4}{\mu_2^2},$$

we see that, as A and a increase, the relation $\frac{\mu_{3,(a)}^2}{\mu_{2,(a)}^3}$ tends to zero and the relation $\frac{\mu_{4,(a)}}{\mu_{2,(a)}^2}$ tends to the limit 3 provided that the relations $\frac{1}{a}\,\frac{\mu_3^2}{\mu_2^3}$ and $\frac{1}{a}\,\frac{\mu_4}{\mu_2^2}$ do tend to zero. If further the relation $\frac{a\,\mu_i}{[a\,\mu_2]^{\frac{i}{2}}}$ tends with increasing a to zero for $i=3,4,5,\ldots,\infty$, then tho relation $\frac{\mu_{2r+1,(a)}}{[\mu_{2,(a)}]^{r+\frac{1}{2}}}$ tends to zero, and the relation

$\frac{\mu_{2r,(\alpha)}}{[\mu_{2,(\alpha)}]^r}$ tends to the limit 1. 3. 5.... (2 r-1) for $r = 1, 2, 3, ..., \infty$ and the law of distribution of the values of $X_{(\alpha)}$ tends to the Gauss-Laplace law, as A and α increase.

§ II

1) Let us distribute the drawings in s groups the first of which contains the first α_1 drawings; the second of which contains the following α_2 drawings and so on. Putting $\alpha_1+\alpha_2+....+\alpha_s=\alpha$ and substituting in the formulae of the chapter IV, § III,2) the values of $m_{1,1}$, $\mu_{1,1}$, $m_{1,1,1}$, $\mu_{1,1,1}$ and so on found above, we find:

$$\boldsymbol{E}\, Z_{(S)} = m_1$$

$$(6) \quad \boldsymbol{E}\, Z^2_{(S)} = m_1^2 + \frac{1}{A\text{-}1}\left[\frac{A}{S}\,\frac{1}{S}\sum_{i=1}^{S}\frac{1}{\alpha_i} - 1\right]\mu_2 \text{ and so on.}$$

$$(7) \quad \boldsymbol{E}\left[Z_{(S)} - m_1\right]^2 =$$

$$= \frac{1}{A\text{-}1}\left[A\,\frac{1}{S^2}\sum_{i=1}^{S}\frac{1}{\alpha_i} - 1\right]\mu_2 = \frac{1}{A\text{-}1}\left[\frac{A}{S}\,\frac{1}{S}\sum_{i=1}^{S}\frac{1}{\alpha_i} - 1\right]\mu_2$$

$$(8) \quad \boldsymbol{E}\left[\frac{1}{S}\sum_{i=1}^{S}\left[Z'_i - Z_{(S)}\right]^2\right] = \frac{A}{A\text{-}1}\,\frac{S\text{-}1}{S}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{\alpha_i}\right]\mu_2$$

$$(9) \quad \boldsymbol{E}\,\frac{1}{\alpha}\sum_{i=1}^{S}\alpha_i\left[Z'_i - X_{(\alpha)}\right]^2 = \frac{A}{A\text{-}1}\,\frac{S\text{-}1}{\alpha}\,\mu_2$$

$$\boldsymbol{E}\left\{\frac{1}{S}\sum_{i=1}^{S}\left[Z'_i - Z_{(S)}\right]^2 - \frac{A}{A\text{-}1}\,\frac{S\text{-}1}{S}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{\alpha_i}\right]\mu_2\right\}^2 =$$

$$(10) \quad = \frac{(S\text{-}1)^2}{S^2}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{\alpha_i}\right]^2\left\{\left[\frac{A(A^2-3A+3)}{(A\text{-}1)(A\text{-}2)(A\text{-}3)}\,\mu_2^2 - \right.\right.$$

$$\left.\left. - \frac{A}{(A\text{-}2)\,(A\text{-}3)}\,\mu_4\right] - \frac{A^2}{(A\text{-}1)^2}\,\mu_2^2\right\} +$$

$$+ \frac{(S\text{-}1)^2}{S^3}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{\alpha_i^3}\right]\left\{\frac{A^2\,(A+1)}{(A\text{-}1)(A\text{-}2)\,(A\text{-}3)}\,\mu_4 - \frac{3A^2}{(A\text{-}2)\,(A\text{-}3)}\mu_2^2\right\} +$$

$$+ \frac{2}{S}\left\{\frac{S\text{-}2}{S}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{\alpha_i^2}\right] + \right.$$

$$+\frac{1}{S}\left[\frac{1}{S}\sum_{i=1}^{S}\frac{1}{a_i}\right]^2\left\{\frac{A(A^2-3A+3)}{(A-1)(A-2)(A-3)}\mu_2^2-\frac{A}{(A-2)(A-3)}\mu_4\right\}$$

(11) $$E\left\{\frac{1}{\alpha}\sum_{i=1}^{S}a_i\left[Z'_i-X_{(\alpha)}\right]^2-\frac{A}{A-1}\frac{S-1}{\alpha}\mu_2\right\}^2=$$

$$=\left\{\frac{1}{\alpha^4}\sum_{i=1}^{S}\frac{(\alpha-a_i)^2}{a_i}\right\}\left\{\frac{A^2}{(A-1)(A-2)(A-3)}\left[(A+1)\mu_4-3(A-1)\mu_2^2\right]\right\}+$$

$$+\frac{(S-1)(S+1)}{\alpha^2}\frac{A}{(A-1)(A-2)(A-3)}\left\{(A^2-3A+3)\mu_2^2-(A-1)\mu_4\right\}-$$

$$-\frac{(S-1)^2}{\alpha^2}\frac{A^2}{(A-1)^2}\mu_2^2$$

2) If all groups are equally numerous, we find, putting $a_i=\alpha_o$ and $S\alpha_o=\alpha$ and noting that under these conditions $Z_{(S)}=X_{(\alpha)}$,

$$E\,Z_{(S)}=E\,X_{(\alpha)}=m_1$$

(12) $$E\left[Z_{(S)}-m_1\right]^2=\frac{A-\alpha}{A-1}\frac{1}{\alpha}\mu_2=\left[1-\frac{\alpha-1}{A-1}\right]\frac{1}{\alpha}\mu_2$$

(13) $$E\,\frac{1}{S}\sum_{i=1}^{S}\left[Z'_i-Z_{(S)}\right]^2=\frac{S-1}{\alpha}\frac{A}{A-1}\mu_2$$

$$E\left\{\frac{1}{S}\sum_{i=1}^{S}\left[Z'_i-Z_{(S)}\right]^2-\frac{S-1}{\alpha}\frac{A}{A-1}\mu_2\right\}^2=$$

(14) $$=\frac{(S-1)(S+1)}{\alpha^2}\left\{\frac{A(A^2-3A+3)}{(A-1)(A-2)(A-3)}\mu_2^2-\frac{A}{(A-2)(A-3)}\mu_4\right\}-$$

$$-\frac{(S-1)^2}{\alpha^2}\frac{A^2}{(A-1)^2}\mu_2^2+\frac{(S-1)^2}{\alpha^3}\left\{\frac{A^2(A+1)}{(A-1)(A-2)(A-3)}\mu_4-\frac{3A^2}{(A-2)(A-3)}\mu_2^2\right\}$$

3) The calculation of μ_2 from the empirical data can proceed in different ways. Considering all drawings as a whole, we find:

(15) $$\mu_2=E\frac{A-1}{A}\frac{1}{\alpha-1}\sum_{i=1}^{\alpha}\left[X'_i-X_{(\alpha)}\right]^2,$$

and

(16) $$E\left\{\frac{A-1}{A}\frac{1}{\alpha-1}\sum_{i=1}^{\alpha}\left[X'_i-X_{(\alpha)}\right]^2-\mu_2\right\}^2=$$

$$= \frac{(A-1)(A-\alpha)}{\alpha(\alpha-1)A(A-2)(A-3)}\left\{\left[\alpha A - A - \alpha - 1\right]\mu_4 - \frac{\alpha A^2-3A^2+6A-3\alpha-3}{A-1}\mu_2^2\right\}$$

or, for A and α sufficiently great numbers of the same order, approximately:

$$E\left[\frac{1}{\alpha}\sum_{i=1}^{\alpha}\left[X'_i - X_{(\alpha)}\right]^2 - \mu_2\right]^2 = \frac{A-\alpha}{A\alpha}\left[\mu_4-\mu_2^2\right]=\left[\frac{1}{\alpha}-\frac{1}{A}\right]\left[\mu_4-\mu_2^2\right]$$

If the drawings are distributed in s groups, we find, on the other hand,

(17) $$\mu_2 = E\frac{1}{\frac{1}{s}\sum\limits_{i=1}^{s}\frac{1}{\alpha_i}}\frac{A-1}{A}\frac{1}{S-1}\sum_{i=1}^{s}\left[Z'_i - Z_{(S)}\right]^2$$

(18) $$\mu_2 = E\frac{A-1}{A}\frac{1}{S-1}\sum_{i=1}^{s}\alpha_i\left[Z'_i - X_{(\alpha)}\right]^2$$

$$E\left\{\frac{1}{\frac{1}{s}\sum\limits_{i=1}^{s}\frac{1}{\alpha_i}}\frac{A-1}{A}\frac{1}{S-1}\sum_{i=1}^{s}\left[Z'_i - Z_{(s)}\right]^2 - \mu_2\right\}^2 =$$

(19) $$= \frac{A^2-3}{A(A-2)(A-3)}\mu_2^2 - \frac{(A-1)^2}{A(A-2)(A-3)}\mu_4 +$$

$$+\frac{\sum\limits_{i=1}^{s}\frac{1}{\alpha_i^3}}{\left[\sum\limits_{i=1}^{s}\frac{1}{\alpha_i}\right]^2}\left\{\frac{A^2-1}{(A-2)(A-3)}\mu_4 - \frac{3(A-1)^2}{(A-2)(A-3)}\mu_2^2\right\} +$$

$$+\frac{2}{(s-1)^2}\left\{s(s-2)\frac{\sum\limits_{i=1}^{s}\frac{1}{\alpha_i^2}}{\left[\sum\limits_{i=1}^{s}\frac{1}{\alpha_i}\right]^2}+1\right\}\left\{\frac{(A-1)(A^2-3A+3)}{A(A-2)(A-3)}\mu_2^2 - \frac{(A-1)^2}{A(A-2)(A-3)}\mu_4\right\}$$

$$E\left\{\frac{A-1}{A}\frac{1}{s-1}\sum_{i=1}^{S}\alpha_i\left[Z'_i - X_{(\alpha)}\right]^2 - \mu_2\right\}^2 =$$

(20) $$= \frac{A-1}{(S-1)^2A(A-2)(A-3)}\left\{\left[A(A+1)\left(\frac{1}{\alpha^2}\sum_{i=1}^{S}\frac{(\alpha-\alpha_i)^2}{\alpha_i}\right) - \right.\right.$$

$$\left. -(A-1)(S^2-1)\right]\mu_4 + \left[(A^2-3A+3)(S^2-1) - \right.$$

$$\left.\left. -\frac{A(A-2)(A-3)}{A-1}(S-1)^2-3A(A-1)\left(\frac{1}{\alpha^2}\sum_{i=1}^{S}\frac{(\alpha-\alpha_i)^2}{\alpha_i}\right)\right]\mu_2^2\right\}$$

If all groups are equally numerous, we have:

$$(21)\quad \mu_2 = E\,\frac{A-1}{A}\,\frac{a_0}{S-1}\sum_{i=1}^{S}\left[Z'_i - Z_{(S)}\right]^2$$

$$(22)\quad E\left\{\frac{A-1}{A}\,\frac{a_0}{S-1}\sum_{i=1}^{S}\left[Z'_i - Z_{(S)}\right]^2 - \mu_2\right\}^2 =$$

$$= \frac{A-1}{A(A-2)(A-3)}\left\{\left[\frac{A(A+1)}{a} - (A-1)\frac{S+1}{S-1}\right]\mu_4 + \right.$$

$$\left. + \left[(A^2 - 3A + 3)\frac{S+1}{S-1} - \frac{A(A-2)(A-3)}{A-1} - \frac{3A(A-1)}{a}\right]\mu_2^2\right\}$$

Confronting (22) and (16), we see that, if $\frac{\mu_4}{\mu_2^2} \leqq A-2$, the standard error of the estimation of the value of μ_2 from the empirical data does increase, if the observations are distributed in equally numerous groups. The calculation of μ_2 on both ways remains, notwithstanding that, of great interest, as we can thus check the assumptions which underlie our calculations: if the value of μ_2 found from all observations, as a whole, does deviate from the value of μ_2 found on the other way more than is compatible with their standard errors, it must be admitted that the assumptions are contradicted by the experiment. Following the exemple of Dormoy-Lexis, we can calculate the quotient of the value of μ_2 found from the grouped observations and of the value of μ_2 found from all observations, as a whole. As criterion appears then the approximate equality of the relation

$$\frac{\frac{1}{s-1}\sum\limits_{i=1}^{s}\alpha_i\left[Z'_i - X_{(\alpha)}\right]^2}{\frac{1}{\alpha-1}\sum\limits_{i=1}^{\alpha}\left[X'_i - X_{(\alpha)}\right]^2} \text{ to } 1.$$

Comparing this criterion with the traditional divergency-coefficient of Lexis, we see that the assumption that drawn tickets are not replaced in the urn does not alter the construction of the criterion. It follows herefrom that by means of the calculation of the Lexis'ian divergency-coefficient the «normal stability» - the case of the replaced tickets - cannot be empirically distinguished from the «super-normal stability» which characterizes the special case of correlated observations when

the tickets are not replaced into the urn. This compels to modificate substantially the theory of dispersion of Lexis.*

§ III

1) Suppose that the A tickets are distributed between t urns. Let the i- th urn contain $A^{(i)}$ tickets,- $a_1^{(i)}$ tickets marked with the number ξ_1, $a_2^{(i)}$ tickets marked with the number ξ_2 and so on; and put:

$$\sum_{i=1}^{t} A^{(i)} = A$$

$$\sum_{i=1}^{t} a_j^{(i)} = a_j$$

$$\sum_{j=1}^{k} a_j^{(i)} = A^{(i)}$$

$$\sum_{j=1}^{k} a_j = A.$$

If, in conformity to the notations

$$m_r = \sum_{j=1}^{k} \frac{a_j}{A} \xi_j^r \, , \; \mu_r = \sum_{j=1}^{k} \frac{a_j}{A} \left[\xi_j - m_1 \right]^r,$$

we put

$$m_r^{(i)} = \sum_{j=1}^{k} \frac{a_j^{(i)}}{A^{(i)}} \xi_j^r \, , \; \mu_r^{(i)} = \sum_{j=1}^{k} \frac{a_j^{(i)}}{A^{(i)}} \left[\xi_j - m_1^{(i)} \right]^r,$$

we find:

$$m_r = \frac{1}{A} \sum_{i=1}^{t} A^{(i)} m_r^{(i)}$$

$$(23) \quad \mu_2 = \frac{1}{A} \sum_{i=1}^{t} A^{(i)} \mu_2^{(i)} + \frac{1}{A} \sum_{i=1}^{t} A^{(i)} \left[m_1^{(i)} - m_1 \right]^2$$

$$\mu_r = \frac{1}{A} \sum_{i=1}^{t} A^{(i)} \mu_r^{(i)} + \sum_{h=1}^{r-2} C_r^h \frac{1}{A} \sum_{i=1}^{t} A^{(i)} \mu_{r-h}^{(i)} \left[m_1^{(i)} - m_1 \right]^h +$$

(*) Cp. A. Tschuprow, *Ist die normale Stabilität empirisch nachweisbar? (Zur Kritik der Lexis' schen Dispersionstheorie)*. «Nordisk Statistisk Tidskrift» (in press).

$$+\frac{1}{A}\sum_{i=1}^{t} A^{(i)}\left[m_1^{(i)}-m_1\right]^r$$

2) Suppose that from the first urn $\alpha^{(1)}$ tickets are drawn and that the tickets which happen to be drawn are not replaced in the urn; that from the second urn $\alpha^{(2)}$ tickets are drawn and so on. Putting $\alpha^{(1)}+\alpha^{(2)}+\cdots+\alpha^{(t)}=\alpha$ and noting that the drawings from the i—th urn are independent from the drawings from the j — th urn, we find for the i — th urn:

$$\mu_{1,1}^{(i)}=-\frac{1}{(A^{(i)}-1)}\mu_2^{(i)}$$

$$\mu_{1,1,1}^{(i)}=\frac{2}{(A^{(i)}-1)(A^{(i)}-2)}\mu_3^{(i)}$$

$$\mu_{2,1}^{(i)}=-\frac{1}{(A^{(i)}-1)}\mu_3^{(i)}$$

$$\mu_{1,1,1,1}^{(i)}=-\frac{6}{(A^{(i)}-1)(A^{(i)}-2)(A^{(i)}-3)}\mu_4^{(i)}+\frac{3A^{(i)}}{(A^{(i)}-1)(A^{(i)}-2)(A^{(i)}-3)}\left[\mu_2^{(i)}\right]^2$$

$$\mu_{2,1,1}^{(i)}=\frac{2}{(A^{(i)}-1)(A^{(i)}-2)}\mu_4^{(i)}-\frac{A^{(i)}}{(A^{(i)}-1)(A^{(i)}-2)}\left[\mu_2^{(i)}\right]^2$$

$$\mu_{2,2}^{(i)}=-\frac{1}{A^{(i)}-1}\mu_2^2+\frac{A^{(i)}}{(A^{(i)}-1)}\left[\mu_2^{(i)}\right]^2$$

$$\mu_{3,1}^{(i)}=-\frac{1}{(A^{(i)}-1)}\mu_4^{(i)}$$

3) Let us denote by $X_h^{(i)}$ the number marked on the ticket which happens to be drawn from the i- th urn at the h- th drawing and by $X_{(\alpha^{(i)})}$ — the average of all numbers drawn from the i- th urn. We have then:

$$X_{(\alpha^{(i)})}=\frac{1}{\alpha^{(i)}}\sum_{h=1}^{\alpha^{(i)}} X_h^{(i)}$$

(24) $$\mathbf{E}\, X_{(\alpha^{(i)})}=m_1^{(i)}$$

$$\mathbf{E}\, X^2_{(\alpha^{(i)})}=\left[m_1^{(i)}\right]^2+\frac{A^{(i)}-\alpha^{(i)}}{(A^{(i)}-1)\,\alpha^{(i)}}\mu_2^{(i)}$$

$$\mathbf{E}\left[X_{(\alpha^{(i)})}-m_1^{(i)}\right]^2=\frac{A^{(i)}-\alpha^{(i)}}{(A^{(i)}-1)\alpha^{(i)}}\mu_2^{(i)}$$

(25) $$\mathbf{E}\left[X_{(\alpha^{(i)})}-m_1^{(i)}\right]^3=\frac{(A^{(i)}-\alpha^{(i)})(A^{(i)}-2\alpha^{(i)})}{(A^{(i)}-1)(A^{(i)}-2)\,[\alpha^{(i)}]^2}\mu_3^{(i)}$$

$$\boldsymbol{E}\left[X_{(\alpha^{(i)})}-m_1^{(i)}\right]^4=\frac{(A^{(i)}-\alpha^{(i)})}{[\alpha^{(i)}]^3}\left\{\frac{[(A^{(i)}-2\alpha^{(i)})(A^{(i)}-3\alpha^{(i)})-A^{(i)}(\alpha^{(i)}-1)]}{(A^{(i)}-1)(A^{(i)}-2)(A^{(i)}-3)}\mu_4^{(i)}+\right.$$
$$\left.+\frac{3A^{(i)}(\alpha^{(i)}-1)(A^{(i)}-\alpha^{(i)}-1)}{(A^{(i)}-1)(A^{(i)}-2)(A^{(i)}-3)}\left[\mu_2^{(i)}\right]^2\right\}$$

$$(26)\qquad \boldsymbol{E}\,\frac{1}{\alpha^{(i)}}\sum_{h=1}^{\alpha^{(i)}}\left[X_h^{(i)}-X_{\alpha^{(i)}}\right]^2=\frac{A^{(i)}}{(A^{(i)}-1)}\,\frac{(\alpha^{(i)}-1)}{\alpha^{(i)}}\,\mu_2^{(i)}$$

$$\boldsymbol{E}\left\{\frac{1}{\alpha^{(i)}}\sum_{h=1}^{\alpha^{(i)}}\left[X_h^{(i)}-X_{\alpha^{(i)}}\right]^2\right\}^2=$$

$$(27)\qquad =\frac{(\alpha^{(i)}-1)\,A^{(i)}\,(A^{(i)}-\alpha^{(i)})\,[\alpha^{(i)}\,A^{(i)}-A^{(i)}-\alpha^{(i)}-1]}{[\alpha^{(i)}]^3\,(A^{(i)}-1)\,(A^{(i)}-2)\,(A^{(i)}-3)}\;\mu_4^{(i)}\;+$$
$$+\frac{(\alpha^{(i)}-1)A^{(i)}\{\alpha^{(i)}(\alpha^{(i)}+1)[A^{(i)}]^2-3(A^{(i)}-1)([\alpha^{(i)}]^2+A^{(i)}\alpha^{(i)}+\alpha^{(i)}-1)\}}{[\alpha^{(i)}]^3\,(A^{(i)}-1)(A^{(i)}-2)(A^{(i)}-3)}\left[\mu_2^{(i)}\right]^2$$

Putting

$$X_{(t)}=\frac{1}{t}\sum_{i=1}^{t}X_{(\alpha^{(i)})},$$

we find:

$$(28)\qquad \boldsymbol{E}\,X_{(t)}=\frac{1}{t}\sum_{i=1}^{t}m_1^{(i)}$$

$$\boldsymbol{E}\,X_{(t)}^2=\left[\frac{1}{t}\sum_{i=1}^{t}m_1^{(i)}\right]^2+\frac{1}{t^2}\sum_{i=1}^{t}\frac{A^{(i)}-\alpha^{(i)}}{(A^{(i)}-1)\,\alpha^{(i)}}\,\mu_2^{(i)}$$

$$(29)\qquad \boldsymbol{E}\left[X_{(t)}-\frac{1}{t}\sum_{i=1}^{t}m_1^{(i)}\right]^2=\frac{1}{t^2}\sum_{i=1}^{t}\frac{A^{(i)}-\alpha^{(i)}}{(A^{(i)}-1)\,\alpha^{(i)}}\,\mu_2^{(i)}$$

Putting

$$X_{(t,A)}=\frac{1}{A}\sum_{i=1}^{t}A^{(i)}\,X_{\alpha^{(i)}}=m_1+\frac{1}{A}\sum_{i=1}^{t}A^{(i)}\left[X_{\alpha^{(i)}}-m_1^{(i)}\right]$$

we find:

$$(30)\qquad \boldsymbol{E}\;X_{(t,A)}=m_1$$

$$\boldsymbol{E}\,X_{(t,A)}^2=m_1^2+\frac{1}{A^2}\sum_{i=1}^{t}\frac{[A^{(i)}]^2(A^{(i)}-\alpha^{(i)})}{(A^{(i)}-1)\alpha^{(i)}}\,\mu_2^{(i)}$$

$$\boldsymbol{E}\left[X_{(t,A)}-m_1\right]^2=\frac{1}{A^2}\sum_{i=1}^{t}\frac{[A^{(i)}]^2(A^{(i)}-\alpha^{(i)})}{(A^{(i)}-1)\,\alpha^{(i)}}\,\mu_2^{(i)}$$

$$(31)\qquad \boldsymbol{E}\left[X_{(t,A)}-m_1\right]^3=\frac{1}{A^3}\sum_{i=1}^{t}\frac{[A^{(i)}]^3(A^{(i)}-\alpha^{(i)})(A^{(i)}-2\alpha^{(i)})}{(A^{(i)}-1)(A^{(i)}-2)[\alpha^{(i)}]^2}\,\mu_3^{(i)}$$

$$E\left[X_{(t,A)}-m_1\right]^4=$$

$$=\frac{1}{A^4}\sum_{i=1}^{t}\frac{[A^{(i)}]^4(A^{(i)}-a^{(i)})[(A^{(i)}-2a^{(i)})(A^{(i)}-3a^{(i)})-A^{(i)}(a^{(i)}-1)]}{[a^{(i)}]^3(A^{(i)}-1)(A^{(i)}-2)(A^{(i)}-3)}\mu_4^{(i)}+$$

$$+\frac{3}{A^4}\sum_{i=1}^{t}\frac{[A^{(i)}]^4(A^{(i)}-a^{(i)})[2a^{(i)}(A^{(i)}-a^{(i)})(2A^{(i)}-3)-A^{(i)}(A^{(i)}-1)^2]}{[a^{(i)}]^3(A^{(i)}-1)^2(A^{(i)}-2)(A^{(i)}-3)}\left[\mu_2^{(i)}\right]^2+$$

$$+3\left\{\frac{1}{A^2}\sum_{i=1}^{t}\frac{[A^{(i)}]^2(A^{(i)}-a^{(i)})}{(A^{(i)}-1)a^{(i)}}\mu_2^{(i)}\right\}^2$$

Putting

$$X_{(t,a)}=\frac{1}{a}\sum_{i=1}^{t}a^{(i)}X_{(a(i))},$$

we find:

(32) $$E\,X_{(t,a)}=\frac{1}{a}\sum_{i=1}^{t}(a^{(i)})\,m_1^{(i)}$$

$$E\,X^2_{(t,a)}=\left[\frac{1}{a}\sum_{i=1}^{t}a^{(i)}m_1^{(i)}\right]^2+\frac{1}{a^2}\sum_{i=1}^{t}\frac{a^{(i)}(A^{(i)}-a^{(i)})}{(A^{(i)}-1)}\mu_2^{(i)}$$

(33) $$E\left[X_{(t,a)}-\frac{1}{a}\sum_{i=1}^{t}a^{(i)}m_1^{(i)}\right]^2=\frac{1}{a^2}\sum_{i=1}^{t}\frac{a^{(i)}(A^{(i)}-a^{(i)})}{(A^{(i)}-1)}\mu_2^{(i)}$$

4) From (30) and (31) we find:

$$E\,X_{(t,A)}=m_1$$

$$\sigma^2_{X_{(t,a)}}=E\left]X_{(t,A)}-m_1\right]^2=\frac{1}{A^2}\sum_{i=1}^{t}\frac{[A^{(i)}]^2(A^{(i)}-a^{(i)})}{(A^{(i)}-1)a^{(i)}}\mu_2^{(i)}$$

If all a tickets had been drawn in the same way from an urn containing all A tickets, we would find, denoting by X_f the number marked on the ticket which happens to be drawn at the f- th drawing,

$$E\frac{1}{a}\sum_{f=1}^{a}X_f=m_1$$

$$E\left[\frac{1}{a}\sum_{f=1}^{a}X_f-m_1\right]^2=\frac{A-a}{(A-1)\,a}\mu_2$$

If the grouping of the observations shall diminish the standard error of the empirical estimation of the value of m_1, the necessary and sufficient condition thereto is, consequently,

$$\frac{\dfrac{1}{A^2}\sum\limits_{i=1}^{t}\dfrac{[A^{(i)}]^2\,(A^{(i)}-\alpha^{(i)})}{(A^{(i)}-1)\,\alpha^{(i)}}\,\mu_2^{(i)}}{\dfrac{A-\alpha}{(A-1)\alpha}\,\mu_2} < 1$$

or (cp. above (23))

$$\frac{\alpha\,(A-1)\sum\limits_{i=1}^{t}\dfrac{[A^{(i)}]^2\,(A^{(i)}-\alpha^{(i)})}{(A^{(i)}-1)\,\alpha^{(i)}}\,\mu_2^{(i)}}{A\,(A-\alpha)\left\{\sum\limits_{i=1}^{t}A^{(i)}\,\mu_2^{(i)}+\sum\limits_{i=1}^{t}A^{(i)}\left[m_1^{(i)}-m_1\right]^2\right\}} < 1$$

If all quantities $m_1^{(i)}$ are equal, this condition can not be fulfilled. The grouping of the observations is then accompanied by diminished precision of the estimation of the value of m_1 provided that the experiment is performed in such manner that the tickets are not replaced in the urn. On the contrary, in the case when the quantities $m_1^{(i)}$ differ considerably, the grouping of the observations can be of great use. Suppose, for instance, that all $A^{(i)}$ are equal, that all $\mu_2^{(i)}$ are equal and equal to $\mu_2^{(o)}$ and that all $\alpha^{(i)}$ are also equal; the quotient of the standard error of m_1 found from grouped observations and of the standard error of m_1 found from ungrouped observations

is then reduced to $\dfrac{\frac{A-1}{A-t}\,\mu_2^{(o)}}{\mu_2^{(o)}+\frac{1}{t}\sum\limits_{i=1}^{t}\left[m_1^{(i)}-m_1\right]^2}$ and can be much

less than 1, if the dispersion of the quantities $m_1^{(i)}$ is great compared to $\mu_2^{(o)}$.

Noting that, if $\sum\limits_{i=1}^{t} Z_i = 1$, $\sum\limits_{i=1}^{t}\dfrac{K_i}{Z_i}$ assumes its smallest value,

when $Z_i = \dfrac{\sqrt{K_i}}{\sum\limits_{i=1}^{t}\sqrt{K_i}}$, we see that the most profitable value of

$\alpha^{(i)}$ which gives the smallest standard error of the empirical estimation of the value of m_1 is

$$\alpha^{(i)} = \frac{\alpha \sqrt{\dfrac{[A^{(i)}]^3 \mu_2^{(i)}}{A^{(i)}-1}}}{\sum\limits_{i=1}^{t} \sqrt{\dfrac{[A^{(i)}]^3 \mu_2^{(i)}}{A^{(i)}-1}}}$$

The quantity

$$\frac{1}{A^2} \sum_{i=1}^{t} \frac{[A^{(i)}]^2 (A^{(i)} - \alpha^{(i)})}{(A^{(i)}-1)\, \alpha^{(i)}} \mu_2^{(i)}$$

is then reduced to its minimal value and becomes equal to

$$\frac{1}{A^2} \left\{ \frac{1}{\alpha} \left[\sum_{i=1}^{t} \sqrt{\frac{[A^{(i)}]^3 \mu_2^{(i)}}{(A^{(i)}-1)}} \right]^2 - \sum_{i=1}^{t} \frac{[A^{(i)}]^2 \mu_2^{(i)}}{(A^{(i)}-1)} \right\}.$$

If the quantities $A^{(i)}$ are sufficiently great, we can put, without appreciable disadvantage

$$\alpha^{(i)} = \alpha \frac{A^{(i)} \sqrt{\mu_2^{(i)}}}{\sum\limits_{i=1}^{t} A^{(i)} \sqrt{\mu_2^{(i)}}}$$

5) Noting that

$$\sum_{i=1}^{t} \left[X_{(\alpha^{(i)})} - X_{(t)} \right]^2 = \sum_{i=1}^{t} X^2_{(\alpha^{(i)})} - t\, X^2_{(t)},$$

we find:

$$E \frac{1}{t-1} \sum_{i=1}^{t} \left[X_{(\alpha^{(i)})} - X_{(t)} \right]^2 =$$

$$= \frac{1}{t} \sum_{i=1}^{t} \frac{(A^{(i)} - \alpha^{(i)})}{(A^{(i)}-1)\alpha^{(i)}} \mu_2^{(i)} + \frac{1}{t-1} \sum_{i=1}^{t} \left[m_1^{(i)} - \frac{1}{t} \sum_{i=1}^{t} m_1^{(i)} \right]^2$$

whereas (see above (23))

$$\mu_2 = \frac{1}{A} \sum_{i=1}^{t} A^{(i)} \mu_2^{(i)} + \frac{1}{A} \sum_{i=1}^{t} A^{(i)} \left[m_1^{(i)} - m_1 \right]^2 .$$

Thus, even in the case when all $m_1^{(i)}$ are equal, the value of μ_2 can not be estimated on the basis of the empirical data by means of the calulation of $\frac{1}{t-1} \sum\limits_{i=1}^{t} \left[X_{(\alpha^{(i)})} - X_{(t)} \right]^2$. We must proceed on a more entangled scheme.

Noting that

$$E \frac{1}{\alpha^{(i)}} \sum_{h=1}^{\alpha^{(i)}} \left[X_h^{(i)} - X_{(\alpha^{(i)})} \right]^2 = \frac{(\alpha^{(i)}-1) A^{(i)}}{\alpha^{(i)} (A^{(i)}-1)} \mu_2^{(i)},$$

we find:

$$(34) \quad \mu_2^{(i)} = E\frac{(A^{(i)}-1)}{A^{(i)}} \cdot \frac{1}{(\alpha^{(i)}-1)} \sum_{h=1}^{\alpha(i)} \left[X_h^{(i)} - X_{(\alpha^{(i)})} \right]^2.$$

On the other hand

$$(35) \quad E\frac{1}{A}\sum_{i=1}^{t} A^{(i)} \left[X_{(\alpha^{(i)})} - X_{(t,A)} \right]^2 = E\frac{1}{A}\sum_{i=1}^{t} A^{(i)} X^2_{(\alpha^{(i)})} - EX^2_{(t,A)} =$$

$$= \frac{1}{A}\sum_{i=1}^{t} A^{(i)} \left[m_1^{(i)} - m_1 \right]^2 + \frac{1}{A}\sum_{i=1}^{t} \frac{A^{(i)}(A^{(i)}-\alpha^{(i)})}{(A^{(i)}-1)\alpha^{(i)}} \mu_2^{(i)} -$$

$$- \frac{1}{A^2} \sum_{i=1}^{t} \frac{[A^{(i)}]^2 (A^{(i)}-\alpha^{(i)})}{(A^{(i)}-1)\alpha^{(i)}} \mu_2^{(i)}.$$

Substituting

$$\frac{1}{A} \sum_{i=1}^{t} A^{(i)} \left[m_1^{(i)} - m_1 \right]^2 = \mu_2 - \frac{1}{A} \sum_{i=1}^{t} A^{(i)} \mu_2^{(i)},$$

we obtain finally:

$$(36) \qquad \mu_2 = E \Bigg\{ \frac{1}{A} \sum_{i=1}^{t} A^{(i)} \left[X_{(\alpha(i))} - X_{(t,A)} \right]^2 +$$

$$+ \frac{1}{A^2} \sum_{i=1}^{t} A^{(i)} \left[(A-1)\alpha^{(i)} - A + A^{(i)} \right] \frac{1}{\alpha^{(i)}(\alpha^{(i)}-1)} \sum_{h=1}^{\alpha(i)} \left[X_h^{(i)} - X_{(\alpha(i))} \right]^2 \Bigg\}$$

CHAPTER VI

§ I

Suppose that the variables X_1, X_2, ..., X_N can assume only two values - 1 and 0 - and that an event A depends on them in such manner that it happens when one of the variables assumes the value 1 and it fails when the variable assumes the value 0. Let us denote by H the frequency of the event A at N observations; by p_h and $q_h = 1 - p_h$ -the probabilities of the values 1 and 0 at the h-th observation; by p_{h_1,h_2} - the probability that the event A happens at the h_1-th and at the h_2- th observations; by p_{h_1,h_2,h_3} - the probability that A happens at the h_1 - th, at the h_2 - th and at the h_3 - th observations and so on; and put:

$$p_0 = \frac{1}{N} \sum_{h=1}^{N} p_h$$

Noting that the average of the observed values of the variables does coincide under these conditions with the frequency of A, we find (cp. chapter I (9)):

$$\boldsymbol{E}\, H = p_o$$

$$\boldsymbol{E}\left[H-p_o\right]^2 = \frac{1}{N^2}\sum_{h=1}^{N} p_h\, q_h + \frac{1}{N^2}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(p_{h_1,h_2} - p_{h_1}p_{h_2}\right)^{(*)} =$$

$$= \frac{1}{N^2}\sum_{h=1}^{N} p_h q_h + \frac{2}{N^2}\sum_{i=1}^{N-1}\sum_{j=1}^{N-i}\left(p_{i,i+j} - p_i\, p_{i+j}\right)$$

$$\boldsymbol{E}\left[H - p_o\right]^3 = \frac{1}{N^3}\sum_{h=1}^{N} p_h q_h\left[q_h - p_h\right] +$$

$$+ \frac{3}{N^3}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(q_{h_1} - p_{h_1}\right)\left(p_{h_1,h_2} - p_{h_1}p_{h_2}\right) +$$

$$+ \frac{1}{N^3}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\sum_{h_3 \neq h_2 \neq h_1}\left(p_{h_1,h_2,h_3} - 3\, p_{h_1} p_{h_2,h_3} + 2 p_{h_1}\, p_{h_2} p_{h_3}\right)$$

$$\boldsymbol{E}\left[H\text{-}p_o\right]^4 = \frac{1}{N^4}\sum_{h=1}^{N} p_h\, q_h\left(q_h^3 + p_h^3\right) +$$

$$+ \frac{3}{N^4}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(q_{h_1} - p_{h_1}\right)\left(q_{h_2} - p_{h_2}\right)\left(p_{h_1,h_2} - p_{h_1}\, p_{h_2}\right) +$$

$$+ \frac{3}{N^4}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1} p_{h_1}\, q_{h_1} p_{h_2}\, q_{h_2} + \frac{4}{N^4}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(q_{h_1}^3 + p_{h_1}^3\right)\left(p_{h_1,h_2} - p_{h_1} p_{h_2}\right) +$$

$$(1) \qquad + \frac{6}{N^4}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\sum_{h_3 \neq h_2 \neq h_1}\left(q_{h_1} - p_{h_1}\right)\left(p_{h_1,h_2,h_3} - p_{h_2}\, p_{h_1,h_3} -\right.$$

$$\left. - p_{h_3} p_{h_1,h_2} + p_{h_1} p_{h_2} p_{h_3}\right) + \frac{6}{N^4}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\sum_{h_3 \neq h_2 \neq h_1} p_{h_1}^2\left(p_{h_2,h_3} - p_{h_2}\, p_{h_3}\right) +$$

$$+ \frac{1}{N^4}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\sum_{h_3 \neq h_2 \neq h_1}\sum_{h_4 \neq h_3 \neq h_2 \neq h_1}\left(p_{h_1,h_2,h_3,h_4} - 4 p_{h_1}\, p_{h_2,h_3,h_4} +\right.$$

$$\left. + 6\, p_{h_1} p_{h_2} p_{h_3,h_4} - 3\, p_{h_1} p_{h_2} p_{h_3} p_{h_4}\right)$$

(*) cp. G. Bohlmann, *Die Grundbegriffe der Wahrscheinlichkeitsrechnung in ihrer Anwendung auf die Lebensversicherung* « Atti del IV Congresso internazionale dei matematici » vol. III; Roma 1909, p. 262.

Noting that the difference $X'_i - X_{(N)}$ can take under these conditions only the values $1 - H$ and $- H$ and that it assumes the value $1 - H$ NH times and the value $- H$ $N\text{-}NH$ times, we find (cp. chapter III (1), (2) and (8)):

$$v'_{[2;N]} = \frac{1}{N}\sum_{i=1}^{N}\left[X'_i - X_{(N)}\right]^2 = H(1\text{-}H)$$

$$(2)\quad v'_{[3;N]} = H(1\text{-}H)(1\text{-}2H)$$

$$v'_{[4;N]} = H(1\text{-}H)\left[1\text{-}3\,H\,(1\text{-}H)\right]$$

and

$$(3)\quad v_{[2;N]} = \mathbf{E}\,H(1\text{-}H) = \frac{1}{N}\sum_{h=1}^{N}\left[p_h - p_o\right]^2 +$$

$$+\frac{N\text{-}1}{N}\left\{\frac{1}{N}\sum_{h=1}^{N} p_h q_h - \frac{1}{N(N\text{-}1)}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(p_{h_1,h_2} - p_{h_1}p_{h_2}\right)\right\}$$

$$v_{[3;N]} = \frac{1}{N}\sum_{h=1}^{N}\left[p_h - p_o\right]^3 + \frac{(N\text{-}1)(N\text{-}2)}{N^2}\left\{\frac{1}{N}\sum_{h=1}^{N} p_h q_h\left(q_h - p_h\right) -\right.$$

$$-\frac{3}{N(N\text{-}1)}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(q_{h_1} - p_{h_1}\right)\left(p_{h_1,h_2} - p_{h_1}p_{h_2}\right) +$$

$$+\frac{2}{N(N\text{-}1)(N\text{-}2)}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\sum_{h_3 \neq h_2 \neq h_1}\left(p_{h_1,h_2,h_3} - 3\,p_{h_1}p_{h_2,h_3} +\right.$$

$$\left.\left. + 2\,p_{h_1}p_{h_2}p_{h_3}\right)\right\} +$$

$$(4)\qquad + \frac{3(N\text{-}2)}{N}\frac{1}{N}\sum_{h=1}^{N} p_h q_h\left[p_h - p_o\right] -$$

$$-\frac{6(N\text{-}1)}{N}\frac{1}{N(N\text{-}1)}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(p_{h_1,h_2} - p_{h_1}p_{h_2}\right)\left[p_{h_1} - p_o\right] =$$

$$= \frac{1}{N}\sum_{h=1}^{N}\left[p_h - p_o\right]^3 + \frac{(N\text{-}1)(N\text{-}2)}{N^3}\sum_{h=1}^{N} p_h q_h\left(q_h - p_h\right) -$$

$$-\frac{3(N\text{-}2)}{N^3}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(q_{h_1} - p_{h_1}\right)\left(p_{h_1,h_2} - p_{h_1}p_{h_2}\right) +$$

$$+\frac{2}{N^3}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\sum_{h_3 \neq h_2 \neq h_1}\left(p_{h_1,h_2,h_3} - 3p_{h_1}p_{h_2,h_3} + 2p_{h_1}p_{h_2}p_{h_3}\right) +$$

$$+\frac{3(N-2)}{N^2}\sum_{h=1}^{N} p_h q_h \left[p_h - q_o\right] - \frac{6}{N^2}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(p_{h_1,h_2} - p_{h_1}p_{h_2}\right)\left[p_{h_1} - p_o\right]$$

$$v_{[4;N]} = \frac{1}{N}\sum_{h=1}^{N}\left[p_h - p_o\right]^4 + \frac{(N-1)(N-2)(N-3)}{N^4}\sum_{h=1}^{N} p_h q_h\left(q_h^3 + p_h^3\right) -$$

$$-\frac{4(N-2)(N-3)}{N^4}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(q_{h_1}^3 + p_{h_1}^3\right)\left(p_{h_1,h_2} - p_{h_1}p_{h_2}\right) +$$

$$+\frac{6(N-3)}{N^4}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\sum_{h_3 \neq h_2 \neq h_1}\left(q_{h_1} - p_{h_1}\right)\left(p_{h_1,h_2,h_3} - p_{h_2}p_{h_1,h_3} -\right.$$

$$\left. - p_{h_3}p_{h_1,h_2} + p_{h_1}p_{h_2}p_{h_3}\right) + \frac{6(N-3)}{N^4}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\sum_{h_3 \neq h_2 \neq h_1}\left(q_{h_1} - p_{h_1}\right)\left(p_{h_2,h_3} -\right.$$

$$\left. - p_{h_2}p_{h_3}\right) - \frac{3}{N^4}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\sum_{h_3 \neq h_2 \neq h_1}\sum_{h_4 \neq h_3 \neq h_2 \neq h_1}\left(p_{h_1,h_2,h_3,h_4} -\right.$$

$$\left. - 4p_{h_1}p_{h_2,h_3,h_4} + 6p_{h_1}p_{h_2}p_{h_3,h_4} - 3p_{h_1}p_{h_2}p_{h_3}p_{h_4}\right) +$$

$$(5) \qquad + \frac{(N-1)(2N-3)}{N^4}\sum_{h=1}^{N} p_h q_h\left(q_h^3 + p_h^3\right) -$$

$$-\frac{4(2N-3)}{N^4}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(q_{h_1}^3 + p_{h_1}^3\right)\left(p_{h_1,h_2} - p_{h_1}p_{h_2}\right) +$$

$$+\frac{3(2N-3)}{N^4}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1} p_{h_1}q_{h_1}p_{h_2}q_{h_2} +$$

$$+\frac{3(2N-3)}{N^4}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(q_{h_1} - p_{h_1}\right)\left(q_{h_2} - p_{h_2}\right)\left(p_{h_1,h_2} - p_{h_1}p_{h_2}\right) +$$

$$+\frac{6(N-2)}{N^3}\sum_{h=1}^{N} p_h q_h\left[p_h - p_o\right]^2 + \frac{6}{N^3}\left[\sum_{h=1}^{N} p_h q_h\right]\left[\sum_{h=1}^{N}\left[p_h - p_o\right]^2 -\right.$$

$$-\frac{12}{N^3(N-1)}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(p_{h_1,h_2} - p_{h_1}p_{h_2}\right)\left[p_{h_1} - p_o\right]^2 +$$

$$+\frac{6}{N^3}\left[\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(p_{h_1,h_2} - p_{h_1}p_{h_2}\right)\right]\left[\sum_{h=1}^{N}\left[p_h - p_o\right]^2\right] +$$

$$+\frac{4(N^2 - 3N + 3)}{N^4}\sum_{h=1}^{N} p_h q_h\left(q_h - p_h\right)\left[p_h - p_o\right] -$$

$$-\frac{12(N-2)}{N^3}\sum_{h_1=1}^{N}\sum_{h_2 \neq h_1}\left(q_{h_1} - p_{h_1}\right)\left(p_{h_1,h_2} - p_{h_1}p_{h_2}\right)\left[p_{h_1} - p_o\right] +$$

$$+\frac{12}{N^3}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}\left(q_{h_2}-p_{h_2}\right)\left(p_{h_1,h_2}-p_{h_1}p_{h_2}\right)\left[p_{h_1}-p_o\right]+$$

$$+\frac{12}{N^3}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}\sum_{h_3\neq h_2\neq h_1}\left(p_{h_1,h_2,h_3}-p_{h_1}p_{h_2,h_3}-p_{h_2}p_{h_1,h_3}-p_{h_3}p_{h_1,h_2}+\right.$$

$$\left.+\,2\,p_{h_1}p_{h_2}p_{h_3}\right)\left[p_{h_1}-p_o\right]$$

$$\sigma^2_{v'_{[2,N]}}=\frac{(N-1)^2}{N^4}\sum_{h=1}^{N}p_h\,q_h\left(q_h^3+p_h^3\right)-$$

$$-\frac{4(N-1)}{N^4}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}\left(q_{h_1}^3+p_{h_1}^3\right)\left(p_{h_1,h_2}-p_{h_1}p_{h_2}\right)+$$

$$+\frac{3(N-1)}{N^4}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}\left(q_{h_1}-p_{h_1}\right)\left(q_{h_2}-p_{h_2}\right)\left(p_{h_1,h_2}-p_{h_1}p_{h_2}\right)+$$

$$(6)\qquad +\frac{3(N-1)}{N^4}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}p_{h_1}q_{h_1}p_{h_2}q_{h_2}-$$

$$-\left\{\frac{N-1}{N^2}\sum_{h=1}^{N}p_h q_h-\frac{1}{N^2}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}\left(p_{h_1,h_2}-p_{h_1}p_{h_2}\right)\right\}^2+$$

$$+\frac{(N-2)(N-3)}{N^4}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}\left(q_{h_1}-p_{h_1}\right)\left(q_{h_2}-p_{h_2}\right)\left(p_{h_1,h_2}-p_{h_1}p_{h_2}\right)+$$

$$+\frac{(N-2)(N-3)}{N^4}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}p_{h_1}q_{h_1}p_{h_2}q_{h_2}-$$

$$-\frac{2(N-3)}{N^4}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}\sum_{h_3\neq h_2\neq h_1}\left(q_{h_1}-p_{h_1}\right)\left(p_{h_1,h_2,h_3}-p_{h_2}p_{h_1,h_3}-\right.$$

$$\left.-\,p_{h_3}p_{h_1,h_2}+p_{h_1}p_{h_2}p_{h_3}\right)-$$

$$-\frac{2(N-3)}{N^4}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}\sum_{h_3\neq h_2\neq h_1}p_{h_1}^2\left(p_{h_2,h_3}-p_{h_2}p_{h_3}\right)+$$

$$+\frac{1}{N^4}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}\sum_{h_3\neq h_2\neq h_1}\sum_{h_4\neq h_3\neq h_2\neq h_1}\left(p_{h_1,h_2,h_3,h_4}-4p_{h_1}p_{h_2,h_3,h_4}+\right.$$

$$\left.+\,6\,p_{h_1}p_{h_2}p_{h_3,h_4}-3p_{h_1}p_{h_2}p_{h_3}p_{h_4}\right)+$$

$$+\frac{4(N-1)}{N^3}\sum_{h=1}^{N}p_h q_h\left(q_h-p_h\right)\left[p_h-p_o\right]+\frac{4}{N^2}\sum_{h=1}^{N}p_h q_h\left[p_h-p_o\right]^2+$$

$$+\frac{4}{N^2}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}\left(p_{h_1,h_2}-p_{h_1}p_{h_2}\right)\left[p_{h_1}-p_o\right]\left[p_{h_2}-p_o\right]+$$

$$+\frac{4(N-1)}{N^3}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}\left(q_{h_1}-p_{h_1}\right)\left(p_{h_1,h_2}-p_{h_1}p_{h_2}\right)\left[p_{h_2}-p_o\right]-$$

$$-\frac{8}{N^3}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}\left(q_{h_1}-p_{h_1}\right)\left(p_{h_1,h_2}-p_{h_1}p_{h_2}\right)\left[p_{h_1}-p_o\right]-$$

$$-\frac{4}{N^3}\sum_{h_1=1}^{N}\sum_{h_2\neq h_1}\sum_{h_3\neq h_2\neq h_1}\left(p_{h_1,h_2,h_3}-p_{h_1}p_{h_2,h_3}-p_{h_2}p_{h_1,h_3}-p_{h_3}p_{h_1,h_2}+\right.$$

$$\left.+2p_{h_1}p_{h_2}p_{h_3}\right)\left[p_{h_1}-p_o\right]$$

§ II

1) Let us put in the formulae of the chapter V $K=2$, $\xi_1=1$, $\xi_2=0$, $\frac{a_1}{A}=p$, $1-p=q$. The average $-X_{\alpha)}$ — does then coincide with the frequency of the ticket marked with 1 and the quantities μ_2, μ_3 and μ_4 assume the values:

$$\mu_2=p\,q$$

$$\mu_3=p\,q\,(q-p)$$

$$\mu_4=p\,q\,(1-3\,p\,q)\,.$$

Hence:

$$\mu_{2,(\alpha)}=\frac{A-\alpha}{A-1}\,\frac{1}{\alpha}\,p\,q$$

(7) $$\mu_{3,(\alpha)}=\frac{(A-\alpha)(A-2\alpha)}{(A-1)\,(A-2)}\,\frac{1}{\alpha^2}\,p\,q\,(q-p)$$

$$\mu_{4,(\alpha)}=\frac{(A-\alpha)}{\alpha^3(A-1)(A-2)(A-3)}\,p\,q\left\{\left[(A-2\alpha)(A-3\alpha)-A(\alpha-1)\right]+\right.$$

$$\left.+\,3\left[A(\alpha-1)(A-\alpha)-(A-2\alpha)(A-3\alpha)\right]p\,q\right\}(*)$$

$$v_{[2,\alpha]}=E\,\frac{1}{\alpha}\sum_{i=1}^{\alpha}\left[X_i-X_{(\alpha)}\right]^2=\frac{\alpha-1}{\alpha}\,\frac{A}{A-1}\,p\,q$$

(*) Cp. « Biometrika » vol. V, p. 174.

$$(8) \qquad v_{[3.a]} = \frac{(a-1)(a-2)}{a^2} \frac{A^2}{(A-1)(A-2)} p\, q \left(q-p\right)$$

$$v_{[4,a]} = \frac{(a-1)(a-2)(a-3)}{a^3} \frac{A}{(A-1)(A-2)(A-3)} pq \left[(A^2-2A+3)-3A^2pq\right] +$$

$$+ \frac{(a-1)(2a-3)}{a^3} \frac{A}{A-1} pq$$

$$(9) \qquad \sigma^2_{v'_{[2,a]}} = \frac{(a-1)A(A-a)}{a^3(A-1)(A-2)(A-3)} p\, q \left\{[Aa-A-a-1] - \right.$$

$$\left. - \frac{2A\,[a(2A-3)-3(A-1)]}{A-1}\, p\, q \right\}$$

2) In the same way, as above, we could easily obtain from chapter V, (1) the values of $m_{2,(a)}$, $m_{3,(a)}$ and $m_{4,(a)}$; but I prefer to give the general formula for $m_{r,(a)}$.

The probability to draw x white balls and a-x black balls from an urn which contains a white balls and A-a black balls is, notoriously, equal to $C_a^x \frac{a^{[-x]}\,[A-a]^{[-(a-x)]}}{A^{[-a]}}$

Denoting by W the number of the white balls which happen to be drawn at N drawings, we have, consequently,

$$\boldsymbol{E}\, W^r = \sum_{x=0}^{a} C_a^x \frac{a^{[-x]}[A-a]^{[-(a-x)]}}{A^{[-a]}} X^r$$

On the other hand, noting that

$$\sum_{i=1}^{s} C_s^i\, m^{[-i]}\, n^{[-(s-i)]} = \left[m+n\right]^{[-s]},$$

we find:

$$\boldsymbol{E}\, W^{[-r]} = \sum_{x=0}^{a} C_a^x \frac{a^{[-x]}[A-a]^{[-(a-x)]}}{A^{[-a]}} X^{[-r]} =$$

$$= \frac{1}{A^{[-a]}} \sum_{x=0}^{a} \frac{a(a-1)\cdots(a-x+1)a(a-1)\cdots(a-x+1)[A-a]^{[-(a-x)]}x(x-1)\cdots(x-r+1)}{1.2.3\ldots.\ X} =$$

$$= \frac{a^{[-r]}a^{[-r]}}{A^{[-a]}} \sum_{x-r=0}^{x-r=a-r} C_{a-r}^{x-r}\, [a-r]^{[-(x-r)]}\, [A-a]^{[-(a-x)]} =$$

$$= \frac{a^{[-r]}a^{[-r]}[A-r]^{[-(a-r)]}}{A^{[-a]}} = \frac{a^{[-r]}a^{[-r]}}{A^{[-r]}}$$

Hence (cp. « *On the math. exp.* », Part I, Introduction (2)) (*)

$$\boldsymbol{E}\, W^r = \boldsymbol{E} \sum_{i=1}^{r} \alpha_{r,i}\, W^{[-i]} = \sum_{i=1}^{r} \alpha_{r,i}\, \frac{a^{[-i]}}{A^{[-i]}}\, \alpha^{[-i]} =$$

$$(10) \qquad = \sum_{j=1}^{r} \alpha^j \sum_{i=0}^{r-j} (-1)^i\, \alpha_{r,j+i}\, \beta_{j+i,i}\, \frac{a^{[-(j+i)]}}{A^{[-(j+i)]}}$$

Putting $\frac{a}{A} = p,\ 1 - p = q,$

we obtain:

$$m_{r,(\alpha)} = \boldsymbol{E} \frac{1}{\alpha^r}\, W^r =$$

$$= \sum_{h=0}^{r-1} \frac{1}{\alpha^h} \sum_{i=0}^{h} (-1)^i\, \alpha_{r,r-h+i}\, \beta_{r-h+i,i}\, \frac{[Ap]^{[-(r-h+i)]}}{A^{[-(r-h+i)]}} =$$

$$(11) \qquad = \frac{[Ap]^{[-r]}}{A^{[-r]}} + \frac{1}{\alpha} \left\{ \alpha_{r,r-1} \frac{[Ap]^{[-(r-1)]}}{A^{[-(r-1)]}} - \beta_{r,1} \frac{[Ap]^{[-r]}}{A^{[-r]}} \right\} + \cdots +$$

$$+ \frac{1}{\alpha^{r-1}} \sum_{i=0}^{r-1} (-1)^i\, \alpha_{r,1+i}\beta_{1+i,i}\, \frac{[Ap]^{[-(1+i)]}}{A^{[-(1+i)]}}$$

Putting

$$\alpha_{r,j+i}\beta_{j+i,i} - \alpha_{r,j+i-1}\beta_{j+i-1,i-1} + \cdots + (-1)^i \alpha_{r,j}\beta_{j,0} = D_{r,j,i},$$

we find from (11) after some transformations

$$\mu_{r,(\alpha)} = \boldsymbol{E} \frac{1}{\alpha^r} \left[W - \alpha p \right]^r =$$

$$= \frac{1}{\alpha^{r-1}}\, pq\, \frac{A}{A-1} \left\{ \frac{-(r-1)}{A}\, \alpha^{r-1} \sum_{h=0}^{r-2} (1)^h\, C_{r-2}^h\, \frac{a^h}{A^h} \frac{[a-1]^{[-(r-2-h)]}}{[A-2]^{[-(r-2-h)]}} + \right.$$

$$(12) \quad \left. + \sum_{j=0}^{r-2} \alpha^j \sum_{h=0}^{j} \sum_{i=0}^{r-2-j} (-1)^{h+i}\, C_r^h\, D_{r-h,j+1-h,i}\, \frac{a^h}{A^h} \frac{[a-1]^{[-(j-h+i)]}}{[A-2]^{[-(j-h+i)]}} \right\} =$$

$$= pq\, \frac{A}{A-1} \left\{ - \frac{r-1}{A} \sum_{h=0}^{r-2} (-1)^h\, C_{r-2}^h\, \frac{a^h}{A^h} \frac{[a-1]^{[-(r-2-h)]}}{[A-2]^{[-(r-2-h)]}} + \right.$$

$$+ \frac{1}{\alpha} \left[\frac{r(r-1)}{2} \sum_{h=0}^{r-2} (-1)^h\, C_{r-2}^h\, \frac{a^h}{A^h} \frac{[a-1]^{[-(r-2-h)]}}{[A-2]^{[-(r-2-h)]}} \right] +$$

$$+ \sum_{j=1}^{r-3} \frac{1}{\alpha^{r-1-j}} \sum_{h=0}^{j} \sum_{i=0}^{r-2-j} (-1)^{h+i}\, C_r^h\, D_{r-h,j+1-h,i}\, \frac{a^h}{A^h} \frac{[a-1]^{[-(j-h+i)]}}{[A-2]^{[-(j-h+i)]}} +$$

$$\left. + \frac{1}{\alpha^{r-1}} \sum_{i=0}^{r-2} (-1)^i\, [i+1]!\, \alpha_{r-1,i+1}\, \frac{[a-1]^{[-i]}}{[A-2]^{[-i]}} \right\}$$

If r be put equal to 2,3,4, we get again the formulae (7).

(*) « Biometrika, vol. XII, p. 142 »

Zusammenfassung

Durch Markoff's Untersuchungen über kettenartig verbundene Versuche, sowie durch die schöne Abhandlung, welche von G. Bohlmann dem vierten internationalen Mathematikerkongress vorgelegt wurde, ist die wahrscheinlichkeitsrechnerische Grundlegung der statistischen Theorie auf neue Wege gewiesen worden, an deren Ausbau sowohl die allgemeine Theorie der Statistik, wie die statistisch verankerten Spezialwissenschaften in hohem Masse interessiert sind. Namentlich hat sich in der statistischen Physik der Gedanke, die Annahme der gegenseitigen Unabhängigkeit der Versuche fallen zu lassen, — in der Gestalt des von Smoluchowski unabhängig von den Mathematikern geformten Begriffs der «Wahrscheinlichkeitsnachwirkung» — als fruchtbar erwiesen.

Das Charakteristische an der Vorgangsweise von Markoff besteht darin, dass an die Stelle der Annahme der Unabhängigkeit der Versuche keine spezialisierten Annahmen in Bezug auf die Art ihrer Verbundenheit eingeführt werden, sondern die allgemeine Formel für den mittleren Fehler des Durchschnittes der empirischen Werte einer zufälligen Variablen abgeleitet wird, von der dann die weitere Behandlung der betrachteten Einzelfälle auszugehen hat. In gleicher Weise wird in den ersten drei Kapiteln der vorliegenden Studie für den Fall von beliebig verbundenen Versuchen das ganze System der Formeln entwickelt, welche die Grundlage für die statistische Behandlung einer zufälligen Variablen bilden. Diesen Formeln liegt nur eine Voraussetzung zu Grunde, — nämlich, dass an einer Variablen, welche bei jedem Versuche verschiedene Werte mit bestimmten Wahrscheinlichkeiten annehmen kann, eine Anzahl von Versuchen ausgeführt wird, welche eine entsprechende Anzahl von empirisch-zufälligen Werten der Variablen liefern. Weitere Voraussetzungen werden nicht gemacht, — weder in Bezug auf die Form der Verteilungsgesetze der Werte der Variablen bei den einzelnen Versuchen, noch in Bezug auf die Art der Verbundenheit der Versuche. Aus den allgemeinen Formeln lassen sich durch geeignete Spezifikationen Formeln für beliebige Einzelfälle ableiten. Im Kap. V werden z. B. aus ihnen die Formeln

für das Schema der «nicht-zurückgelegten Nummern» deduziert; im Kap. VI wird zunächst die BOHLMANN' sche Formel aus der MARKOFF' schen, als ein Spezialfall davon, erhalten und dann, als Ergänzung zur BOHLMANN' schen Untersuchung, das ganze System von Formeln für den von ihm behandelten Fall aus den allgemeinen Formeln der ersten drei Kapitel in gleicher Weise abgeleitet.

Die allgemeinen «voraussetzungslosen» Formeln der ersten drei Kapitel gestatten, die Grenzen der voraussetzungslosen Empirie bei der Behandlung des statistischen Materials scharf abzustecken. Es zeigt sich, dass eine voraussetzungslose Betrachtung der empirischzufälligen Werte, welche durch die statistische Beobachtung geliefert werden, zu keinen weitgehenden Schlüssen zu führen vermag. Der arithmetische Durchschnitt der empirischen Werte behält zwar unter allen Umständen den Charakter eines Näherungswertes der durchschnittlichen mathematischen Erwartung der Variablen. Die Präzision dieses Näherungswertes bleibt jedoch im allgemeinen Falle gänzlich unbestimmt, — nicht einmal darauf darf man sich ohne weiteres verlassen, dass der mittlere Fehler durch die Vermehrung der Beobachtungen beliebig herabgesetzt werden kann. Es sind solche Arten von Verbundenheit zwischen den Versuchen denkbar, bei welchen das Gesetz und das statistische Gebot der grossen Zahlen nicht mehr gelten.

Die Unergiebigkeit der voraussetzungslosen Deutung der empirischen Werte wirft die Fragen auf, unter welchen Voraussetzungen sichere Schlüsse in Bezug auf die ihnen zu Grunde liegenden apriorischen Grössen statthaft sind und wie man sich in jedem Einzelfalle vergewissern kann, ob die einen oder die anderen Voraussetzungen tatsächlich zutreffen. Auf die prinzipielle Bedeutung dieser Fragen für die statistische Theorie wird in der «Introduction» kurz hingewiesen.

Das im fünften Kapitel behandelte Schema der nicht-zurückgelegten Nummern entspricht vielfach den tatsächlichen Voraussetzungen, von welchen die Behandlung des statistischen Materials auszugehen hat, genauer, als das Schema der gegenseitigen Unabhängigkeit der Versuche, weshalb die Engländer mit Vorliebe bei ihren stochastischen Konstruktionen davon auszugehen pflegen. Eine besondere Bedeutung gewinnt dieses Schema dadurch, dass das an Ansehen von Jahr zu Jahr gewinnende repräsentative Verfahren in der Praxis meistens eine ihm angepasste Form annimmt. Die stochastischen Grundlagen des auf diese Art gestalteten repräsentativen Verfahrens werden im fünften Kapitel eingehend analysiert.

Namentlich wird in § IV die Frage nach den Bedingungen untersucht, unter welchen es von Vorteil für die Schätzung des Mittelwertes ist, die zu untersuchende Masse in Teilmassen zu zerlegen, bevor die zu messenden Einheiten ausgelost werden. An sich ist die Zerlegung in Teilmassen bei dieser Modalität des repräsentativen Verfahrens von Nachteil, da der mittlere Fehler des Näherungswertes der gesuchten Grösse durch die Zerlegung vergrössert wird. Die Dispersion der Mittelwerte der Teilmassen kann jedoch den Nachteil mehr als aufwiegen, falls sie hinreichend gross ist im Vergleich zu der durchschnittlichen Dispersion der Teilmassen.

J. W. Bispham, R. E., O. B. E.

An experimental determination of the distribution of the partial correlation coefficient in samples of thirty.

Samples from a highly correlated universe.

1. *Introductory.*

In a previous paper(1) the author described an experimental determination of the *partial* correlation coefficients in samples of thirty from an uncorrelated universe. The distribution was found to be not significantly different from the theoretical(2) or observed distributions of *total* correlation coefficients for samples of the same size.

The present paper extends the investigation to samples from a very highly correlated universe. In the equation expressing the partial correlation coefficient in terms of the appropriate total correlation coefficients, viz.

$$_1r_{23} = \frac{r_{23} - r_{12}\, r_{13}}{\sqrt{(1 - r_{12}^2)\,(1 - r_{13}^2)}} \quad \text{(i)}$$

the denominator of the right-hand-side becomes vanishingly small as r_{12}, or r_{13}, approaches unity. In this case, $_1r_{23}$ becomes a quotient of two small quantities and there would appear, *prima facie*, to be some grounds for expecting higher variability.

In order to test this point closely values of ρ_{12} and ρ_{13} as high as .975 were dealt with. Such values of ρ gave occasional samples in which r was greater than .99 so that the test was a very thorough one.

---

(1) « Proc. Roy. Soc. » A. vol 97. 1920 p. 219.
(2) « Biometrika », vol. 11, p. 328, et seq.

2. *Nature of sampled population.*

The sampled population was an artificial one obtained as follows:

a. *Case of population for which ρ_{12} and ρ_{13} are of the order .85.*

Two sets of fifteen bone discs or «counters» had each marked upon them one of the numbers 1 to 15; so that when the thirty counters were placed together in a bag there were present two representatives of each of the numbers 1 to 15. From the thirty counters ten were drawn at random, and as they were drawn the numbers upon them were written down. The ten numbers were added together and the sum so obtained formed an individual value of the first attribute (A) to be correlated. The relation of the second and third attributes to the first is best indicated algebraically, thus:

$$\text{First attribute (A)} = (a_1 + a_2 + a_3 \,....\, a_9 + a_{10})$$
$$\text{Second attribute (B)} = (a_x + a_2 + a_3 \,....\, a_9 + a_{10})$$
$$\text{Third attribute} = \text{(C)} = (a_1 + a_2 + a_3 \,....\, a_9 + a_y)$$

In the second attribute, itself the sum of ten items a_x replaces the a_1 of the first attribute. Similarly the third attribute is obtained by replacing a_{10} of the first attribute by a_y.

a_x and a_y were themselves items similar to the rest, determined by returning all the counters to the container, shaking up and making two new draws. The first two items drawn were taken to be a_x and a_y respectively.

It will be seen therefore that every individual value dealt with in the correlations represented a sum of ten items, the first and second attributes (A) and (B) had 9 common elements and one differing at random. Similary (A) and (C) had nine common elements; (B) and (C), on the other hand, had 8 common elements and 2 differing at random.

A theoretical examination of this case leads to the prediction that the values of ρ_{12}, ρ_{13}, ρ_{23}, respectively are .862, .862 and .740.

This case was not pursued in great detail. About 3000 individuals were obtained and correlated in samples of thirty. The *partial* correlation coefficients were found to be approximately grouped about zero as mean, but did not give a significantly greater dispersion than those obtained from an uncorrelated universe. It was therefore decided to investigate cases of still greater skewness.

b. *Case of a population for which ρ_{12} and ρ_{13} are of the order .95.*

In this case every individual attribute was the sum of thirty component items each item being drawn independently as descri-

bed above. The first attribute (A) and the second (B) had 29 elements in common; so too had (A) and (C).

Similarly (B) and (C) had 28 common elements and two differing at random.

In all, the three attributes of each of 6,000 individuals were obtained and these were correlated in the order obtained in 200 samples of thirty each. The values of r_{12}, r_{13} and r_{23} for the whole six thousand individuals taken as one sample were .9559, .9553 and .9111 respectively. The observed value of the partial correlation based on this sample of 6000 was—.0248. It was theoretically deduced that the values of ρ_{12}, ρ_{13} and ρ_{23} should be .951, .951 and .905 respectively. (see appendix).

c. *Case of a population for which ρ_{12} and ρ_{13} are of the order .975.*

Every individual was a total of 60 items obtained as above. The pairs (A) and (B) and (A) and (C) had 59 common elements and 1 differing at random. (B) and (C) had 58 common and 2 differing at random. The preliminary arithmetic in these cases was, of course, very bulky.

In all the three attributes of each of 3000 individuals were obtained and these were correlated in the order obtained in 100 samples of 30 each. The values r_{12}, r_{13} and r_{23} for the whole 3000 individuals taken as one sample were .9770, .9766 and .9551 respectively. The value of the observed partial correlation based on this sample of 3000 was + .0220. It was theoretically deduced that the values of ρ_{12}, ρ_{13} and ρ_{23} should be .9754, 9754 and .9513 respectively.

Size of samples. In all cases, whether the individual attributes were totals of ten, or thirty or pixty items, they were correlated in sample groups of thirty and all the results described below are concerned with samples of that size.

3. *Comparison of observed distribution of total correlation coefficients with theoretical distribution.*

The theoretical distribution of total correlation coefficients for various sizes of sample and various values of ρ has been treated comprehensively in a co-operative study (1) to which reference has been made above. The particular examples dealt with formed an extremely good test of the theory for cases of extreme skewness. In the paper referred to tables are given of the ordinates of the

(1) « Biometrika », Vol. II. p. 328.

frequency curve of correlation coefficients for values of ρ from 0 to .9 for samples of various sizes up to 400. The particular case of samples of 30 is not dealt with although distributions for values of ρ up to .9 are given for samples of 25 and 50. Further, no frequency distribution for values of ρ greater than .90 was evaluated. The method of interpolation moreover was clearly inadeguate to give even an approximate value of the ordinates so that in the present investigation it was necessary, applying the equations of the co-operative study to calculate *ab initio* the theoretical distribution for the valuess of ρ of in question for samples of 30. Of the various alternative methods of computating the ordinates described the method of «expansions» was the one employed. It was found that successive terms converged with sufficient rapidity to save excessively laborious arithmetic. The working was checked in one case by realculatincg by an alternative method. There was the further check that on the application of graduation formulae the total frequency under the curve was correct.

At first the distributions computed were for the following even values of ρ, namely .99, .98, .975, .97, .95, and .90. These ordinates and frequency distributions are given in Appendix I. They indi cate the extreme steepness of the modal region of these excessively skew curves. It was hoped that the distributions for the values of ρ «most appropriate» to the observed distributions would be obtainable by interpolation from such closely placed values. In practice, this was found not to be the case as a large part of the frequency area was comprised in an extremely narrow strip lying about the mode, the ordinates differing so greatly for small changes in the value of ρ and r that interpolation did not give satisfactory results. In addition to the above values therefore, the distributions were also calculated for the following « most appropriate » values of ρ, namely .977, .955 and .912.

Two main groups are discussed in detail; firstly, two hundred samples of thirty for which ρ_{12} and $\rho_{13} = .955$ and $\rho_{23} = .912$. This will be referred to as Group I. Secondly, one hundred samples of thirty for which ρ_{12} and $\rho_{13} = .977$ and $\rho_{23} = .955$. This will be referred to as Group II. Below are tabulated the « most appropriate » predicted frequency distributions together with the observed frequencies of *total* correlations. At the foot of each column of observed values is given the value of P (the measure of goodness of fit) for the relation to the predicted distribution.

TABLE I. *Total correlation Coefficients.*
Frequency distribution, $n = 30$, $\rho = .977$.

Range of r	Group II r_{12}	(1000 samples of 30) r_{13}	Predicted frequency
(1)	(2)	(3)	(4)
.99 to 1.0	2	1	1.6
.98 to .99	40	37	37.1
.97 to .98	41	41	39.9
96 to .97	15	14	15.2
.95 to .96	1	5	4.3
.94 to .95	1	1	1.1
.93 to .94	0	1	0.3
TOTAL	100	100	100
	$P = .78$	$P = .90$	

TABLE II. *Total Correlation Coefficients.*
Frequency distribution $n = 30$ *and* $\rho = .955$ *in each case.*

Range of r	Group I ($N = 200$)			Group II ($N = 100$)	
	r_{12}	r_{13}	Predicted frequency	r_{23}	Predicted frequency
(1)	(2)	(3)	(4)	(5)	(6)
.99 to 1.0	0	0	.0	.0	0
.98 to .99	3	2	3.8	2	1.9
.97 to .98.	24	27	27.8	12	13.9
.96 to .97	56	48	59.0	28	25.0
.95 to .96	52	50	47.1	20	23.6
.94 to .95	28	34	32.2	21	16.1
93. to 94.	16	19	18.6	10	9.3
.92 to .93	8	11	9.8	3	4.9
.91 to .92	8	4	5.0	3	2.5
.80 to .91	5	5	5.7	1	2.8
TOTAL	200	200	200	100	100
	$P = .76$	$P = .91$		$P = .84$	

TABLE III. *Total Correlation Coefficients.*

Frequency distribution n $= 30$ *and* $\rho = .912$.

Range of r	Group I (200 samples of 30)	
	r_{23}	Predicted frequency
(1)	(2)	(3)
.97 to 1.0	1	.6
.95 to .97	14	14.2
.93 to .95	39	44.1
.91 to .93	52	53.0
.89 to .91	46	40.1
.87 to .89	19	23.9
.85 to .87	15	12.5
.75 to .85	14	11.3
.65 to .75	0	3
TOTAL	200	200
	$P = .82.$	

It will be seen that there is a very striking agreement indeed between the predicted and observed distributions in these cases. The collected values of P, are .78, .90, 76, .91, .84 and .82, so that more than 8 times out of 10 on the average worse agreement would result from the errors of random sampling. The means of the observed distributions are as follovs:

GROUP I. (200 samples)

.9536 ± .0009

.9529 ± .0009

.9073 ± .0016

GROUP II. (100 samples)

.9774 ± .0006

.9758 ± .0006

.9541 ± .0009

4. *Observed distribution of partial correlation coefficients.*

In order to determine whether a partial correlation coefficient obtained in an investigation is significant or not it is of advantage to know what is the probability that it should appear as a result of random sampling from a true zero value. In this investigation therefore, as a first step, it was intended to determine the range of variation of partial correlation coefficients for the case in which ${}_1\rho_{23}$ was zero. In order to obtain a zero value for ${}_1\rho_{23}$ if is necessary that the relation should hold

$$\rho_{23} - \rho_{12} \cdot \rho_{13} = 0 \text{ (ii)}$$

The device by which it was hoped to attain this, namely by dealling with totals of a number of component items drawn at random, has been described above. That this method was successful is indicated by the fact that after the actual distributions and the means of the distributions of total correlations had been obtained and the « most appropriate » values of ρ_{12}, ρ_{13} and ρ_{23} deduced from them, the condition (ii) was fulfilled to the extent shown below.

For Group I.

$$\rho_{12} = \rho_{13} = .977 \quad \text{and} \quad \rho_{23} = .955$$

so that

$$\rho_{23} - \rho_{12}\,\rho_{13} = .000471$$

For Group II

$$\rho_{12} = \rho_{13} = .955 \quad \text{and} \quad \rho_{23} = .912$$

$$\rho_{23} - \rho_{12}\,\rho_{13} = .000025.$$

It is seen therefore that the condition holds that the true values of the partial correlation coefficients range about a mean value not significantly different from zero.

The following table gives a detailed statement of the observed frequency distribution of *partial* correlation coefficient together with the predicted values based on the cooperative study referred to above for total correlation coefficient. A similar table is given in Part I of this investigation for the case of partial correlation coefficients from an uncorrelated universe.

TABLE IV. *Partial Correlation Coefficients*
($n = 30$ in each case)

	Group I (200)		Group II (100)	
(1)	(2)	(3)	(4)	(5)
Range of r	$_1r_{23}$	Predicted frequency	$_1r_{23}$	Predicted frequency
+ .475 to + .525	1	.6	0	.3
+ .425 » + .475	2	1.0	1	.5
+ .375 » + .425	0	2.2	1	1.1
+ .325 » + .375	4	3.8	4	1.9
+ .275 » + .325	7	6.2	2	3.1
+ .225 » + .275	5	9.0	6	4.5
+ .175 » + .225	16	12.2	8	6.1
+ .125 » + .175	16	15.6	7	7.8
+ .075 » + .125	18	18.4	11	9.2
+ .025 » + .075	19	20.2	10	10.1
— .025 » + .025	15	20.8	11	10.4
— .025 » — .075	18	20.2	6	10.1
— .075 » — .125	21	18.4	8	9.2
— .125 » — .175	19	15.6	10	7.8
— .175 » — .225	13	12.2	4	6.1
— .225 » — .275	11	9.0	1	4.5
— .275 » — .325	5	6.2	5	3.1
— .325 » — .375	2	3.8	4	1.9
— .375 » — .425	2	2.2	1	1.1
— .425 » — .475	4	1.0	0	.5
— .475 » — .525	2	.6	0	.3
TOTAL	200	200	100	100

In the case referred to as Group I (200 cases) $\rho_{12} = \rho_{13} = .955$ and $\rho_{23} = .912$. In Group II one hundred samples $\rho_{12} = \rho_{13}$ .977 and $\rho_{23} = .955$.

The above Table is given in more condensed form below on .15 ranges about the mid-values indicated and the goodness of fit is indicated at the feet of the appropriate columns.

TABLE V. *Partial Correlation Coefficients*
(n = 30 in each case)

	Group I (200)		Group II (100)	
(1)	(2)	(3)	(4)	(5)
Mid-Value of r	${}_1r_{23}$	Predicted frequency	${}_1r_{23}$	Predicted frequency
0.60 etc.	0	.4	0	.2
0.45	3	3.7	2	1.8
0.30	16	19.1	12	9.5
0.15	50	46.2	26	23.1
0	52	61.3	27	30.6
— 0.15	53	46.2	22	23.1
— 0.30	18	19.1	10	9.5
— 0.45	8	3.7	1	1.8
— 0.60 etc.	0	.4	0	.2
TOTALS	200	200	100	100
	$P = .50$		$P = .96$	

The observed distributions of *partial* correlation coefficients are very well fitted by theoretical distributions of *total* correlation coefficients. If the dispersions are expressed in summary form by the value of σ a similar result is obtained.

For Group I $\sigma_r = .1944 \pm .007$
Theoretical value $= .1857$

The difference between the observed and theoretical values is only slightly greatar than the probable error and therefore not significant.

For Group II $\sigma_r = .1901 \pm .009$
Theoretical value $= .1857$

The difference in this case is less than half the probable error.

5. *Conclusions.*

It has been shown in Part I of this investigation that for samples of 30 in the case of *partial* correlation coefficients from an uncorrelated universe, that the dispersion observed is not significantly different from that of *total* correlation coefficients derived from samples of the same size. In all 1000 samples were dealt with so that the results have considerable weight. The present paper deals with smaller groups of samples (200 in one case) because of the prohibitively laborious arithmetic involved but the evidence of these groups from a highly correlated universe is to the same effect, namely that the range of dispersion, owing to errors of sampling, of *partial* correlation coefficients whose true value is zero is not significantly different (the samples being of the same size) from either the dispersion of the corresponding *partial* correlation coefficients derived from uncorrelated material, or the dispersion of *total* correlation coefficients whose mean value is zero.

The method of partial correlation is often used in order to avoid the spurious correlation that may be obtained in using indexes; as for example in correlating mortality rates. In using the method of partial correlation in such cases very high *total* correlations have frequently to be used and it was to test the validity of the use of extremely high total correlation coefficients in deducing partial correlation coefficients that the present investigation was undertaken. The results, so far as they go, appear to indicate that no seriously large increase of dispersion in the partial correlation coefficients is likely to be introduced, even when the samples are as small as 30, by making use of total correlation coefficients as high as .975.

In Part I of this paper the practical importance of the results there obtained was indicated. The present investigation indicates that a still further freedom in the use of the partial correlation coefficient is justifiable. The preliminary arithmetical work has been extremely laborious. It has involved the handling of individual « counters », of something of the order of, half a million times and the subsequent summing in groups of ten, thirty or sixty, in order to obtain the individual attributes. This would not have been possible without the generous help of scool children (1), who made

(1) Most of the work done by the school children was supervised by Miss A. Fowler and to her and to Mr M. E. Wilson for a large share in the calculation and checking of the coefficients I am especially indebted.

the drawings, recorded them and summed and independently checked the results.

In conclusion, I owe much to the help and advice of Dr. E. C. Snow in the early stages, and of Dr. M. Grenwood in the later stages of the work.

Appendix

Ordinates of frequency curves

Table VI. $n = 30$. $N = 1,000$

Value of r	$\rho = .99$	$\rho = .98$	$\rho = .977$	$\rho = .975$
(1)	(2)	(3)	(4)	(5)
.998	174.2	.1	.0	.0
.996	18010.4	205.6	21.7	9.5
.994	78952.1	1706.1	565.9	279.2
.992	116595.0	8849.8	3634.5	2027.0
.990	105558.6	22855.8	8783.0	7018.0
.985	35411.2	56201.1	40621.0	31352.4
.98	8142.0	53061.2	51159.2	46783.4
.975	1795.6	33827.8	40643.4	42554.5
.97	4717.7	17719.5	25737.9	30098.1
.965	105.2	8586.7	14503.7	18595.3
.96	28.9	4017.5	7701.2	10672.2
.955	8.6	1863.0	3977.1	5890.8
.95	2.7	868.6	2034.5	3192.2
.945	.9	410.5	1042.0	1719.6
.94	.3	197.4	537.9	916.3
.935	.1	—	280.9	504.2
.93	.0	48.5	148.7	276.5
.92	.0	12.9	43.6	86.2
.91	.0	3.7	13.6	28.3
.90	.0	1.1	4.5	9.8

Ordinates of Frequency Curves.

TABLE VII. $n = 30. \quad N = 1000$

Value of r	$\rho = .955$	$\rho = .950$	$\rho = .912$	$\rho = .900$
(1)	(2)	(3)	(4)	(5)
.995	.1	.1	.0	.0
.99	79.8	27.0	.1	.0
.985	1375.4	629.0	—	—
.98	6169.0	3355.8	41.7	6.2
.975	14010.7	8858.2	—	—
.97	21469.6	17370.5	726.0	273.1
.965	25706.8	20825.8	—	—
.96	26193.2	23512.3	3270.8	529.1
.955	23879.1	23504.8	—	—
.95	20146.0	21545.3	7420.5	4190.5
.945	16085.3	18543.0	—	—
.94	12345.0	15235.5	11311.4	7537.4
.935	9220.6	12094.4	—	—
.93	6728.2	9356.5	13422.5	10359.2
.925	4845.7	7101.4	—	—
.95	3408.6	5313.7	13503.0	11875.3
.915	2451.4	3934.7	—	—
.91	1724.2	2891.8	12127.1	11991.1
.905	1212.6	2114.2	—	—
.90	352.1	1540.4	10056.2	11051.6
.89	—	—	7876.1	9525.3
.88	—	—	5918.9	7810.4
.87	—	—	4316.2	6168.1
.86	—	—	3079.3	4677.8
.85	26.2	65.9	2162.2	3556.3
.80	1.3	3.6	329.8	710.9
.75	.1	.3	48.9	128.2
.70	.0	.0	7.7	23.8
.65	—	—	1.3	4.5
.60	—	—	.2	.9

TABLE VIII. *Frequency distributions*

$n = 30. \quad N = 1000$

Range of r	Frequency for values of ρ at head of columns							
	.990	.980	.997	.975	.955	.950	912	.900
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
.99 to 1.00	541.2	41.4	16.4	10.5	.1	.1		
.98 to .99	430.8	498.0	371.7	298.1	19.1	9.7	3.0	1.0
.97 to .98	26.5	341.3	400.2	411.0	139.1	92.5		
.96 to .97	1.4	92.9	152.9	191.5	250.0	204.5	70.8	35.2
.95 to .96	.1	20.5	42.9	62.7	235.7	229.0		
.94 to .95		4.5	11.4	18.2	161.1	182.7	220.3	148.8
.93 to .94		—	3.0	5.3	92.9	120.1		
.92 to .93		1.3	1.0	1.8	49.0	71.0	265.2	232.6
.91 to .92		—	.3	.5	24.8	39.4		
.90 to .91		.1	.1	.2	12.3	21.2		
.89 to .90							200.7	218.9
.87 to .89							119.6	156.2
.85 to .87			.1	15.4	15.4	28.5	62.6	94.7
.75 to .85						1.3	56.5	108.7
.65 to .75							1.3	3.8
.55 to .65							.0	.1
TOTAL	1000	1000	1000	1000	1000	1000	1000	1000

Tables VI, VII and VII are calculated from the equations of the co-operative study referred to above.

Raymond Pearl

The interrelations of the biometric and experimental methods of acquiring knowledge: with special reference to the problem of the duration of life (1)

I

The problem of the duration of life would seem to transcend in its significance and interest nearly all other problems of phenomenal biology. By that I mean that, short of the existence of life itself, as a cosmic phenomenon, the obvious fact that the duration of that state of organization of matter which is recognized as living is generally finite and limited, rather than infinitely extended, presents a problem of the first importance to the inquiring biologist. Why do not living aggregates of matter persist in that state to a degree of extension of continuity comparable in magnitude to that exhibited by non-living aggregates? And why do some living things live so much longer than others, similar at least in the important respect that they are living? This problem of the duration of life has interested me because it has seemed that in addition to the intrinsic interest and obvious practical bearing of the problem *per se*, it has a much heightened significance by virtue of the considerable *a priori* probability that if we succeed in attaining some reasonable insight into the problem of duration, we shall concurrently get a greatly improved understanding of life itself.

I wish to invite your attention this evening to a consideration of certain aspects of this problem of the duration of life. These aspects will be chiefly methodological rather than phenomenal. In

---

(1) *Papers from the Department of Biometry and Vital Statistics*, School of Hygiene and Public Health, Johns Hopkins University. No. 56.

The substance of a lecture to the Harvey Society of New York, April 8, 1922.

other words I wish to examine with some particularity what methods are, on the one hand, available, and, on the other hand, likely to prove most useful for the elucidation of this problem. I make no apology for occupyng your time with a consideration of methodology, because none is needed. The plentiful exhibition of quinine in malaria may not be immediately the most entertaining thing in the world, but it is indubitably good for the patient. If I correctly understand your President's invitation to take over the management of your case this evening, what he particularly wished me to do was to administer a reasonable dose of methodological quinine.

II

Every thoughtful student of medicine. and of biology generally, has been impressed with the ever-increasing prevalence and importance of the quantitative viewpoint in his science throughout its historical development, but to an impressive and accelerating degree during the last quarter of a century. The fact is obvious, and needs no elaboration. It is usually interpreted as a somewhat isolated phenomenon in the history of the biological sciences, assumed to parallel what is thought to have occurred at a much earlier stage in the development of the physical sciences. I am inclined to the view, however, that it represents, in its more recent developments certainly, only a phase of a much more profound alteration which has been going on in the methodology of sciences generally and inclusively. This alteration has resulted from the invasion of statistical modes of thought and reasoning into all branches of science, whether physical or biological. (1) It is impossible to say precisely when this process began. But it is apparent enough, on the one hand, that prior to CLERK MAXWELL's work the fundamental concepts of physics had been essentially non-statistical in character, and, on the other hand, that the general viewpoint of LINNAEUS and his contemporaries in biology was equally non-statistical. It became apparent to a number of philosophically minded persons, working in different branches of science, about the middle of the last century, that groups or aggregates of things, whether inanimate or animate, had distinctive and important attributes of their own *as groups*, not discoverable or even describa-

(1) Cf. MERZ, J. T. *A History of European Thought in the Nineteenth Century* Vol. II. 1903.

ble in terms of the attributes of any single individual in the group, however typical that individual might be. This was equally true whether the population was composed of gas molecules or of men. Now the distinctive, and at the same time unique, contribution of the statistical method in science is that it furnishes a means of describing accurately and scientifically a group, in terms of the group's attributes rather than in terms of those of any individual in it. (1) These methods have in recent years become highly refined. In the beginning when CLERK MAXWELL began to think about the attributes of populations of molecules, and DARWIN about populations of living things called species, it was essentially with the idea that populations as such have attributes distinct from those of any individual contained in them.

This idea has been extraordinarily fruitful in the last quarter of a century. To it we owe, essentially, such fundamental scientific advances as the kinetic theory of gases, natural selection, Mendelian inheritance, and a host of other major developments in nearly every branch of science. Medicine has been singularly slow to implement to its advantage this perhaps greatest advance of the nineteenth century in philosophically scientific modes of thought. Why this happened seems clear. It is because the great outstanding event of the nineteenth century in medicine was the founding and development of bacteriology. The philosophical effect of this development upon the progress of the science of medicine was scarcely less profound and much less happy than its material effect. The immediately resulting tendency, and one from which medicine is only now beginning to get away again, was somewhat unduly to stress the idea that the science of etiology began and ended and was comprised within the discovery of the bacterial factor involved in each particular disease. In his recently revised and extremely valuable treatise on constitutional pathology BAUER (2) states the case with great clearness in the following words:

« *Diese nichts weniger denn neue Erkenntnis der Multiplizität ätiologischer Faktoren ist nun zu einer Zeit vielfach vergessen oder zum mindesten vernachlässigt worden, in der die Bakteriologie als führende medizinische Disziplin die volle Erledigung von*

---

(1) Cf. PEARL, R. *Modes of Research in Genetics.* New York, (Macmillan) 1915. Chapters II and III.

(2) BAUER, J. *Die Konstitutionelle Disposition zu inneren Krankheiten.* Zweite Auflage. Berlin, 1921.

Problemen vortäuschte, wo nur Teilfragen dieser Probleme gelöst waren. Gegen die ungebührliche Vernachlässigung ätiologisch in gleichem Masse wirksamer Momente zugunsten eines einzelnen aüsseren, bequemerweise fassbaren und ersichtlichen Faktors erfolgte alsbald eine Reaktion, eine Reaktion die bei VERWORN *so weit ging, den Ursachenbegriff nicht nur für die Medizin, sondern überhaupt fallen zu lassen und auszumerzen* ».

It was at once recognized after PASTEUR had paved the way, that a bacterial organism was probably an obligate factor in the causation of, at least, many diseases. To recover this organism from a case of the disease, cultivate it in pure culture, and to reproduce the disease by inoculating the pure culture into a normal animal, and then recover the organism from the animal once more, was some times taken to prove that this organism was the sole cause of the disease. Of course such a procedure logically proves nothing of the sort, but only that the bacterial organism is one factor associated in an obligate manner with the occurrence of the disease in question. There may be, and often are, one or more other factors equally obligate for the production of the disease. This is well illustrated in the case of tuberculosis. Nearly all persons of adult age in cities have at some time or other been definitely infected with tubercle bacilli. But, roughly speaking, at the most only about 1 in 10 ever develops clinical tuberculosis. The reason for this, as is now plain enough in general terms if not in details, is that other conditions are necessary to produce clinical tuberculosis in addition to the presence of the tubercle bacillus (1). Some of these conditions we perceive in a general way to be directly biological, such as the inherited constitution of the host, while some are undoubtedly immediately related to the physical environment. The fact that the precise quantitative relations of the fairly considerable number of factors that are involved in this problem have yet to be worked out, in no way detracts from the main point that something more than the presence of the tubercle bacillus in the body is necessary to cause clinically active tuberculosis.

Methodologically speaking, the intellectual content of bacteriology has been, until very recently, generally non-statistical in

---

(1) Cf. KRAUSE, A. K. *Environmental factors in tuberculosis*, « Amer. Rev. Tbc. » Vol. 4, pp. 713-728. 1920.

Also PEARL, R. *The relative influence of the constitutional factor in the etiology of tuberculosis. Ibid.* pp. 688 - 712.

character. It has worshipped rather exclusively at the shrine of the crucial experiment. The consequence is that we know little as compared with what we shall presently know about such obviously important matters as the specific reactions and relations generally of groups or populations of men as groups, to group or populations of bacteria as groups. We have given a name, epidemiology, to one particular phase of this subject and the leading workers in that field recognize clearly that the phenomena of epidemiology are group phenomena, and that there is only one method of dealing adequately with group phenomena, the statistical.

III

In order to understand clearly just what it is that the statistical method has to offer medicine it will be desirable to examine briefly the various methods which are available for acquiring knowledge of natural phenomena. Broadly speaking there are but two general methods, viz:

1. The descriptive or historical.
2. The experimental.

Some may wonder why the statistical is not included as a third method particularly as a precedent has been set for it, by so distinguished an authority in these matters as the late JOSIAH ROYCE (1).

This might be done if one feels strongly about it, but it seems to me that the statistical method is in the main only one aspect or phase of the descriptive method philosophically considered.

Because of the nature of things, the two methods, description and experiment, have proved in the course of the scientific experience of mankind to be of quite different value in respect of usefulness. Neither can be dispensed with, but the experimental method unquestionably furnishes, when successful, a far deeper and more certain knowledge of the true causal relationships of phenomena than does the descriptive method. Why this must be the case will be apparent if we consider for a moment, in general terms, the nature of phenomenal relations as we know them. I shall with a good deal of diffidence, venture to present a view of the organization of nature, considered philosophically, which is in no way original but which I have found helpful in planning and conducting research.

---

(1) ROYCE, J. *The mechanical, the historical and the statistical.* « Science ». 1914.

Consider a particular phenomenon or event A (Fig. 1). Let this be of any nature *per se* that one pleases. We know, if we have any knowledge about it at all, that its existence, or its magnitude, or its attributes generally, in both a qualitative and a quantitative sense, are influenced and finally determined by a series of other phenomena or events, which generally may be termed causal factors affecting A. These are in part diagramatically represented in

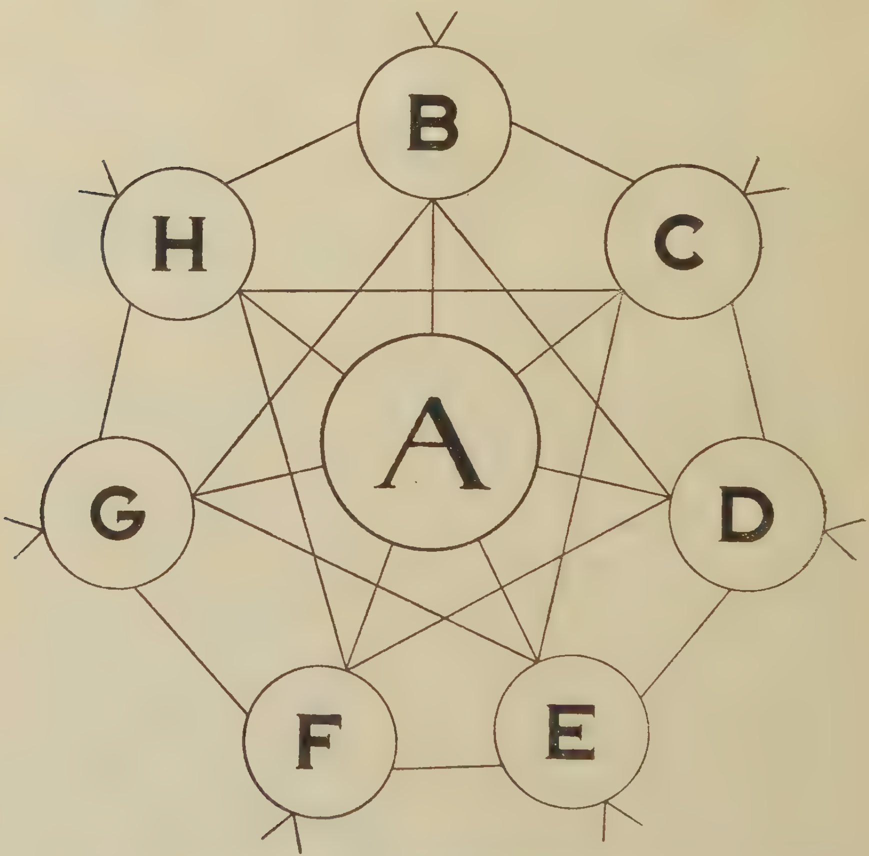


Fig. 1. Hypothetical diagram of cause relationships discussed in text.

the figure by the circles B to H surrounding A. The magnitude of the influence in a quantitative sense of some one (say B) of these causal factors upon A may be, and quite commonly is, overwhelmingly greater than that of any other one, or indeed of all the others put together. When the latter fact is true we say, in common parlance, that B is *the* cause of A, meaning merely that

its influence in determining A's state, condition, or existence so far outweighs all the rest, that these others may for practical purposes be neglected in establishing a basis for the conduct of our lives relative to A. But always if we search the matter minutely, and to a philosophical rather than a purely practical end, it is apparent that in some degree, perhaps extremely minute or remote, C, D, E, and all the rest are involved in the determination of A.

Again the situation may be such that instead of one causal factor B being, in a quantitative sense, overwhelmingly potent in determining A, there may be several factors, say B, E, and H, each about equally influential in determing A's state or existence.

Furthermore it is apparent that since B, C, D, E, F, G and H are themselves phenomena or events, any one of them might equally well replace A in the diagram, and form the center or unknown element for which we sought to find the significant causes. In other words any one of our circles, so far as we know anything about the organization of nature from accumulated scientific experience, is connected with every other one by correlational, « functional, » or causational bonds, represented in the diagram by the connecting lines. Some of the bonds are very long and very tenuous, of course; so slight indeed as not to be measurable or observationally detectable, but philosophically there seems no escape from the conclusion that they must exist. A phenomenon completely isolated, both spatially and temporally from all or any other phenomena, something in short which had neither antecedents or consequences, or associates, is unthinkable, for the simple reason that experience has never presented the human animal with such a phenomenon.

In other words we may conceive the phenomenal universe to be an infinite net work, filling space and time in all directions, of mutually and generally correlated events. (1) The degree of correlation quantitatively considered, between any two of these events may be anything between O and $\pm$ 1, as limits. When the correlation between A and B say approaches closely to the limit + 1, and B in time is always antecedent to A (even by the slightest amount) we say commonly that B is *the* cause of A. And if the correlation between C and A approaches close to the limit O, we

---

(1) Cf. Franklin, W. S. *A much needed change of emphasis in meteorological research.* « Monthly Weather Rev. » Vol. 46, pp. 449-453' October 1918.

are accustomed in the practical usages of life, to say that these two phenomena or events are in no way related. The task of science is, in theory, to measure precisely the value of every one of the correlational bonds. Practically what happens is that we pick out of the great manifold which is nature, a certain event or phenomenon in which for some reason we are interested, and then endeavor in the most scientific manner possible to find out the relative strength of the correlational (or « causational ») bonds between it and a relatively few other events or phenomena, which from *a priori* considerations we think likely to have important relations to it.

To do this science makes use, as has already been said, of the descriptive and the experimental methods. The descriptive method has as its ideal to make a generalization about the correlational or causational bonds between A and B, C, D,... N, by describing carefully what the relationship is observed to be between these phenomena, at some *particular* time or times and at some *particular* point or points in space, the number of times or spaces being finite and generally small. From what is observed in a particular situation, supposed to have come about solely from the action of undisturbed natural causes, the descriptive or historical method draws inferences as to what the causes leading to the situation were. On the other hand, the experimental method goes about the same task in quite a different way. It says, in effect: « I will forcibly control B so that it shall remain *constant* during the course of the experiment, and leave A to vary. In this way, by comparing the behavior of A when B is constant, with its behavior when B is free to vary I shall find just in what degree and in what way B influences A ». So similarly the influence of C can be determined, and by holding both B and C constant their combined influence, and so on up to the limitations imposed by the physical possibilities of experimental control.

The statistical method functions, in the main, as an adjunct and supplement to the descriptive and to the experimental methods: in relation to the former by enabling a more precise and quantitative description of any phenomenon, and by furnishing the only known method of describing a group, as group, in terms of its own attributes rather than in terms of the attributes of the individuals composing it. In relation to the experimental method the statistical calculus makes possible the thing which the experimental method in itself wholly lacks, namely the quantitative

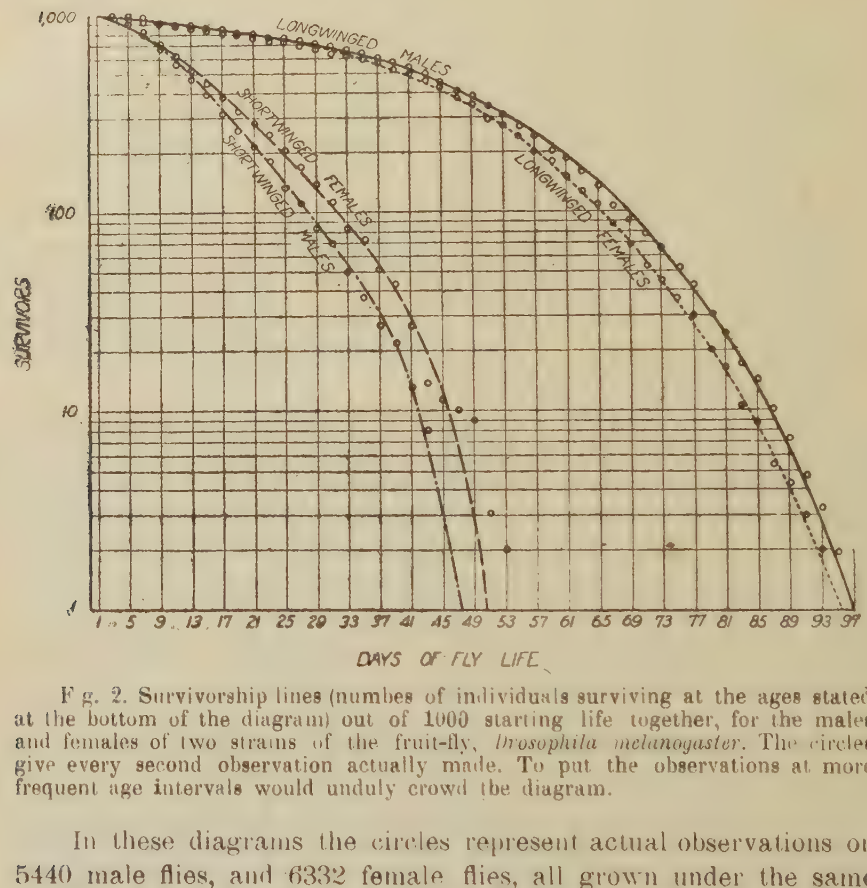
evaluation of the results. For, to go back to our diagram, in order to interpret correctly the behavior of an event or phenomenon *A* while another, *B*, is held constant, as compared with *A*'s behavior when *B* is not held constant, we need always to know what proportion of *A*'s behavior in both cases is purely fortuitous or due to chance. This information can alone be given by the proper application of the statistical calculus. Science knows no other method of correctly evaluating this essential datum. Parenthetically it should be remarked, of course, that the so-called fortuitous or random chance element in *A*'s behavior is presumably merely the element due to the infinite multitude of causational factors remote in time and space acting upon *A*, each separate one to an infinitesimally slight degree, and in infinitely varying directions. All these remote infinitesimal cosmic influences together produce that element in the result, which in common speech we call « chance. » (1).

IV

Let me give briefly some concrete examples of the helpfulness of the statistical method along the lines suggested in what has just been said, first considering the statistical method as the handmaiden of the descriptive method. Duration of life is obviously a quantitative phenomenon. It therefore follows that for its proper analytical investigation one essential requirement will be the most refined method of quantitative description possible. Such a method has been well worked out for this particular problem by the actuaries, because of its important practical bearings upon the conduct of human life. If we are to study with any real penetration the problem of duration of life in a lower organism like *Drosophila* we need to describe the facts about this character with the greatest attainable precision. This means that we must construct life tables for these flies just as the actuary does for human beings.

(1) It will be noted that this is the position of complete determinism. It seems to me that there are two overwhelming objections to any form of indeterminism which is self-consistent or logically structured, at least so far as science is concerned: in the first place there is no sound evidence for it, and in the second place, if adopted as a working basis for the conduct of life, it renders completely nugatory and futile any attempt to arrive at a knowledge of nature by means of scientific investigation. It should in fairness be said, however, that some persons take another view of the matter. Cf. Franklin, W. S. *loc. cit.* and Boussinesq, J. *Conciliation du véritable déterminisme méchanique avec l'existence de la vie et de la liberté morale.* Paris 1878.

This we have done (1). Figure 2 shows one function, the survivorship, for both sexes and two stocks of *Drosophila melanogaster*. The «long-winged» stock is of the wild type and the «shortwinged» stock is a type, carrying in genetically pure form several mutations.



F g. 2. Survivorship lines (numbes of individuals surviving at the ages stated at the bottom of the diagram) out of 1000 starting life together, for the males and females of two strains of the fruit-fly, *Drosophila melanogaster*. The circles give every second observation actually made. To put the observations at more frequent age intervals would unduly crowd the diagram.

In these diagrams the circles represent actual observations on 5440 male flies, and 6332 female flies, all grown under the same constant conditions of temperature, food, etc. The smooth curves represent the ordinates of equations of the type

$$\log 1_x = e\, g^{ax} (a + bx + cx^2 + dx^3)$$

fitted to the observations by the method of least squares. These curves furnish a true scientific description of the facts regarding duration of life in *Drosophila*. They enable exact and quantitative

---

(1) PEARL, H, and PARKER, S. L. *Experimental studies on the duration of life. I. Introductory discussion of the duration of life in* DROSOPHILA. « Amer. Nat. » Vol. 55, pp. 481-509, 1921.

comparisons with other forms in respect of duration of life, and a statement of the attributes of *Drosophila* populations as such in regard to this character.

Thus we can see at once that the law of mortality in *Drosophila* is fundamentally similar to the law of mortality in man. This is shown in Fig. 3.

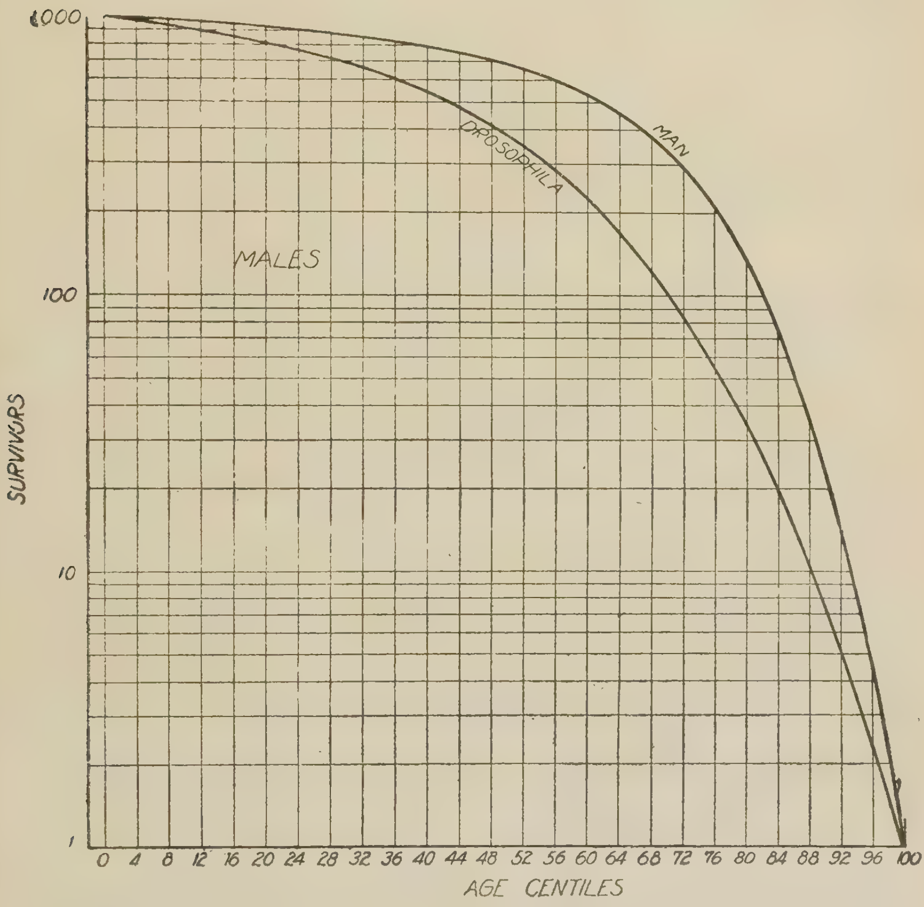


Fig. 3 Comparing the survivorship distributions of *Drosophila* and man (males in both cases) over the equivalent life spans.

Since for reason which need not be gone into here we start our observations on the duration of life of *Drosophila* at the beginning of its life as an imago (upon emergence from the pupal stage of its life history), we leave out the component of the life curve which corresponds to the mortality of infancy and childhood in man.

Upon what basis shall any life table function, say l_x, of the *Drosophila* life table be compared with that of man? The life span of one of these organisms is best measured in days, while that of the other is measured in years. This fact, however, offers no insuperable difficulty to the comparison. What is needed is to superimrose the two curves so that at least two *biologically equivalent* points coincide. The best two points would be the beginning and the end of the life span. But in the case of *Drosophila* our life tables start with the beginning of *imaginal* life only.

I think we can best get at this starting point by putting the human and *Drosophila* l_x curves together as a starting point at the age for each organism *where the instantaneous death rate q_x is a minimum*. In the case of *Drosophila*, I think we are safe in concluding, on the basis of the work of LOEB and NORTHROP as well as from our own observations, that this point is at or very near the beginning of imaginal life. We shall accordingly take *Drosophila* age 1 day as this point. Our life tables show that certainly *after* this time q_x never again has so low a value.

The latest edition of GLOVER's United States Life Tables gives for white males in the original registration states the following values for q_x: for age 11-12 2.28, and for age 12-13, 2.29. We may, therefore, with sufficient accuracy take exactly 12 years as the minimum point, particulary as the l_x figures we shall have to use are tabled as of the *beginning* of the age interval.

For the other end of the life span we may conveniently take the age at which there is left but *one* survivor out of 1,000 starting at age 1 day for *Drosophila* and age 12 years for white males. This age for wild type *Drosophila* is, to the nearest whole figure, 97 days. It appears that in man there is almost exactly one survivor at 98 years out of 1000 starting at age 12. So then we have as biologically equivalent life spans

97 *days* of *Drosophila* life as imago = 86 *years* of human life.

Whence it follows that

1 day of *Drosophila* life =.8866 year of human life and
1 year of human life = 1.1279 days of *Drosophila* life.

We are now in position to make an exact comparison between the life tables of the two organisms. This may be done perhaps most instructively by setting up l_x lines for the two forms on the basis of age in *centiles of the life span*, rather than days

or years. That is to say, the whole comparable life spans of 97 days in *Drosophila* and of 86 years for white males will each be divided into 100 equal parts, and the survivors at the attainment of the beginning of each centile interval will then be computed.

This is done for wild-type (long-winged) *Drosophila* males and male whites in original Registration states in 1910 in Fig. 3, plotted on an arithlog grid. We have, in Fig. 3, for the first time, so far as I am aware, a precise quantitative comparison of the life spans and one of the mathematical functions of the mortality of two different organisms.

It will be noted that:

1. The form of the l_x distributions is fundamentally the same in both of these organisms over the equivalent life spans. Considering the extreme differences in habits of life, structure, physiology, and environmental stresses and strains in the two cases, this is a truly remarkable result. It seems to me to mean that the factors which determine individual longevity, and differences in this character, are biologically deeply rooted, at least as fundamental, apparently, as the factors which determine the specificity in the morphogenesis of organisms, and perhaps even more so. We are accustomed loosely to think that the prime factors in determining human longevity are such things as the infectious diseases, exposure to unfavorable environment, etc. But *Drosophila*, which, so far as is known, has no infectious diseases, and in general meets a set of environmental conditions wholly different, both qualitatively and quantitatively, from those which operate on man, shows fundamentally the same form of distribution of degrees of longevity.

2. When compared exactly, on the basis of comparable life spans, the human being has at every equivalent age a higher relative expectation of life than does *Drosophila*, measured in terms of its own life span in each case.

From this fact the conclusion appears warranted that while the laws of mortality are fundamentally the same *in kind* for *Drosophila* and for man, they differ somewhat quantitatively. There is a temptation to conclude further that the quantitative difference finds its cause in man's own control and amelioration of his environment though sanitation and hygiene. Such a conclusion, however, seems to me not to be strictly warranted, in the light of our present knowledge. There is some suggestion that it is true, from the fact that the progressive change of the human l_x curve in

form during historical times has been in the direction of moving from the form typical of *Drosophila* to that now found for progressive, highly civilized groups of men. But definitive conclusions on the point must await further research.

So much for this example of the application of refined statistical methods to what is at bottom a problem in accurate scientific description. Let us consider another of a wholly different character, on the experimental side. This example is taken from a recent paper by two clinical workers (1) in Germany and deals with the effect upon uric acid excretion of radiating the thymus gland. Though of interest, no great importance appears to attach to the investigation *per se*, but it may well serve as an illustration of method. Six experiments were performed upon as many different subjects. Each experiment consisted essentially of the following procedure. The uric acid output of the patient was measured on one, two or three days immediately prior to the day on which the thymus was radiated. Then on a certain day X-rays were applied to the thymus and the uric acid output on that day determined. Thus for 1, 2 or 3 days following the day of radiation the uric acid output was measured. The results in grouped form are shown in Table I.

TABLE I.

Uric acid output in relation to radiation of thymus.
(Data of Rother and Szegö).

Experiment	Uric acid output		
	Mean of 1, 2, or 3 days before day of radiation	Day of radiation	Mean of 1, 2, or 3 days after radiation
1	0.364 g.	0.388 g.	0.474 g.
2	.353	.386	.515
3	.262	.364	.286
4	.355	.402	.363
5	.273	.642	.539
6	.336	.455	.316
	A	B	C

(1) ROTHER, J. and SZEGÖ, E. *Übür die Beeinflüssung der Harnsäureausscheidung durch Röntgenbestrahlung der Thymusdrüse.* « Zeitschr. f. d. ges. exper. Med. » Bd. 24, pp. 262-269-1921.

The first question which one wishes to answer in experimental work of this kind, where the series of observations or experiments is, for one reason or another, statistically relatively short, is as to whether the differences which appear in the results can be regarded as really significant. Or, put in another way; what is the probability that the differences are purely fortuitous, and without necessary relation to the experimental procedure? In the particular illustration under discussion one wants to know whether the increase in uric acid output on the day of radiation, and on the days following can be regarded as significant, when the whole basis of experience on which it is possible to reason comprises but 6 experiments.

The statistical method enables one to get an exact answer to questions of this sort. I do not propose to bore you with the details of the procedure, but will merely say that a method (1) has been worked out for dealing with these short series. In our particular example, the probability that the mean output of uric acid on the day of radiation (Column B) is really greater than the mean output on the days immediately preceding radiation (Column A) is equal to .959, or the odds are 23 to 1 that the difference is significant. Similarly the probability that the average output of uric acid after radiation of the thymus (Column C) is really greater than before radiation (Column A) is equal to .952, or the odds are 20 to 1 that the difference is significant. This puts the case on a quantitative footing. If one is an optimist he will regard an event which will happen 95 or 96 times in every 100 trials as a certainty. If one is a pessimist he will say that the event is not quite certain but rather highly probable. But in either event, and this is the important consideration, we have a quantitative measure of just precisely how much basis there is for either optimism or pessimism in the premises. And let it be emphasized again that this method, and other similar ones, are just as applicable to important experimental problems as to the perhaps somewhat trivial example I have used for illustrative purposes.

The service which the statistical method can render to medicine, as an adjunct to the descriptive and experimental methods, is of a high order of value. And I would strongly emphasize, that in rendering this service the statistical method, if intelligently used

---

(1) «Student», *The probable error of a mean.* «Biometrika» Vol. 6, pp. 1-25, 1908.

as a helpful adjunct to the descriptive and experimental methods, should not and need not wholly dominate the situation, or by any apparent tendency to do so, work to the confusion and disgust of the non-statistically trained or minded experimentalist. After all, what has just been said as to the way in which the statistical, method can be helpful to the investigator merely means that it is a powerful and trustworthy aid in enabling the experimentalist to draw, on the one hand, correct conclusions, and on the other hand, more and more penetrating conclusions than without it he would be able to attain. Furthermore just because it can do this the statistical method is of the greatest aid in planning experimental work. With a well-trained statistician of vision and imagination at his side, the experimentalist will find his equipment for his own work enormously augmented.

V

In addition to these adjuvant functions which have been mentioned the statistical method offers one wholly unique boon to the investigator. It enables him te get the precise logical equivalent of an experimental result on large classes of material, for which in the inherent nature of things the physical manipulation of the objects in a laboratory experiment is wholly and absolutely impossible. Because of the really great significance of this point I should like to develop it in some detail (1).

The essence of the experimental method, as practiced in the laboratory, and in theory, is that, of the multitude of variables conditioning a phenomenon, as many as possible are, by appropriate methods, held constant while one variable (or at most a very few selected variables) is allowed to vary and the results are noted. One may then deduce the relative significance of the selected variable in determining the phenomenon under observation. One frequently hears about the experiments that Nature makes. Actually the true conditions of an experiment are rarely if ever realized in the course of natural events. It is just because Nature permits manifold and haphazard changes in all variables at the same time, that recourse must be had to the method of experi-

(1) In so doing I shall follow closely a discussion of the matter which I have already given in another place. Cf. PEARL. R. *Influenza Studies* II. «Public Health Reports». Vol. 36. p. 274, 1921.

mental control in the laboratory. What is needed in order to interpret the results, in the experimental sense, and determine the meaning of the manifold and ceaseless changes and variations in the flow of naturally determined events, is a method of picking out of the manifold some selected constant conditions of a series of variables, and then measuring the extent and character of the variations in a single selected variable, whose true relative influence upon the phenomenon it is desired to know, while all these other variables are held constant. If this can be done we shall have realized all the epistemological advantages of the experimental method as practiced in the laboratory, and have freed ourselves at the same time from the limitations which in so many cases inhere in the material itself and make the laboratory type of experimental inquiry impossible. In other words, we shall have let Nature perform the experiment, in the sense of determining the phenomena, in her own way, while we evaluate the results in critically analytical terms of precisely the same sort and meaning as those in which we evaluate the results of a laboratory experiment.

Now exactly this boon is actually afforded in a very wide range of natural phenomena by the method of partial or net correlation, if properly handled. This method enables one, out of a manifold complex of variables operating in an entirely uncontrolled and natural manner, to determine the variation of any selected single variable, or the correlation of any selected pair for constant conditions or values of the other variables in the complex with only the limitation that the regressions shall be linear. I judge that the possibilities of this method are not yet fully grasped by scientific men generally. When they are I believe it will rank as a fundamental method of acquiring knowledge, combining certain of the advantages of the descriptive or historical and of the experimental methods. It seems to me much more effectively to justify ROYCE's (1) eulogy of the statistical method than any of the arguments which he advanced. The best elementary, and at the same time adequate, account of the method is that of YULE (2).

I have recently applied this calculus of partial correlation (3) to an epidemological problem of the first importance, the problem

---

(1) ROYCE, J. *Loc. cit.*

(2) YULE, G. UDNY. *An Introduction to the Theory of Statistics.* Fifth Edition, pp. 228-253, London, 1919.

(3) PEARL, R. *Influenza Studies,*, I - IV incl. «Public Health Reports». Vol. 34 pp. 1743-1783, 1919, and Vol. 36, pp. 273-298, 1921.

of the determination of any factors which had a significant influence in causing the obviously great, but very puzzling variation observed between different cities in this country in respect of the explosiveness of outbreak of epidemic mortality and its total destructiveness in the 1918 pandemic of influenza. By the application of the technique of partial correlation it was possible to demonstrate, with the exactness and precision of a critical laboratory experiment, the following broad facts:

1. Neither the age distribution, nor the proportion of the sexes, nor the density of the population had any significant influence whatever upon either the explosiveness of outbreak of the epidemic mortality or its total destructiveness in the 1918 infiuenza epidemic in American cities.

2. The same was true in respect of the longitude of the city and the rate of its recent growth in population, a factor which in some degree at least is indicative of its industrial activity.

3. On the other hand, the explosiveness of outbreak of the epidemic mortality and its total destructiveness were to an absolutely high and statistically certainly significant degree influenced by the biological make-up of the population of any particular city, as indicated by its normal pre-epidemic average rate of mortality from organic diseases of the heart.

4, The normal death rate of the population from pulmonary tuberculosis had a significant influence upon explosiveness of outbreak, but only doubtfully so upon total destructiveness.

5. The normal death rate from a acute nephritis and Bright's disease definitely influenced explosiveness of outbreak, but was wholly without relation to total destructiveness.

6. Tha normal death rate from typhoid fever was wholly without influence upon either explosiveness or destructiveness. The same was true of cancer.

I shall not discuss these results in detail here, as I wish merely to show by a concrete example how modern statistical methods enable us to get just as clean cut analytical results as experimentation would, in an instance where the possibility of experimental procedure is excluded, from the nature of the case. It is further of interest to see that in this case it has been possible to get at information as to biological relationships, apparently of a very fundamental character, which so far as I am able to see could

not have been reached by any but the statistical method. And the way to further research on the problem of epidemic influenza is clearly indicated. If the normal status of the population in respect of organic disease of the heart was the most potent influence yet demonstrated in determining how severely a community was going to be attacked by the influenza epidemic, then clearly the next thing we want to know is how this relationship was brought about biologically. The necessity for a different type of clinical study of influenza than any so far done is clearly indicated. And further, one wants to know what is the reason or basis for the wide variation among different communities in respect of the normal incidence or prevalence of organic diseases of the heart. Certain phases of this latter problem we are now studying in this laboratory.

VI

The analysis up to this point seems to me to have indicated clearly that the most effective method now available to acquire useful knowledge of the significant causes of dynamic phenomena is by an intelligent combination of the experimental and the statistical modes of research. I wish now to return to the discussion of the main problem of this paper, the problem of the duration of life, and give some concrete illustrations of the effectiveness of this methodological combination.

The problem of duration of life is fundamentally a problem of the causation of variation. What factors are most significant, in a broad sense, in determining why one individual lives substantially longer than another? Since we are dealing here with living things it is at once apparent on purely *a priori* grounds, that any significant factor influencing duration of life must fall into one or the other of the two broad categories, heredity and environment. Or, to put the matter in another way, it will be clear that factors of heredity and factors of environment are both involved, in some unknown proportion or ratio, in determining how long an individual shall live, just so soon as it shall have been demonstrated that this character is inherited at all.

Is duration of life inherited? Unequivocally the answer is affirmative. There can be no doubt on this point. It is impossible to present all the evidence here, but two quite different and independent sets of results will demonstrate the fact. The first is purely

statistical and derived from ALEXANDER GRAHAM BELL's(1) *Study of the Hyde genealogy.* Table 2 shows the extent to which the expectation of human life at birth is increased by virtue solely of having relatively longlived *parents.*

TABLE II.

Showing the influence of a considerable degree of longevity in both father and mother upon the expectation of life of the offspring (After BELL*). In each cell of the table the open figure is the average duration of life of the offspring and the bracketed figure is the number of cases upon which the average is based.*

Father's age at death	Mother's age at death		
	Under 60	60-80	Over 80
Under 60	32.8 years (128)	33.4 years (120)	36.3 years (74)
60-80	35.8 (251)	38.0 (328)	45.0 (172)
Over 80	42.3 (131)	45.5 (206)	52.7 (184)

We see that the longest average duration of life, or expectation of life at birth, was of that group which had both mothers and fathers living to age 80 and over. The average duration of life of these persons was 52.7 years. Contrast this with the average duration of life of those whose parents both died under 60 years of age, where the figure is 32.8 years. In other words, it added almost exactly 20 years to the average life of the first group of people to have long-lived parents, instead of parents dying under age 60. In each column of the table the average duration of life advances as we proceed from top to bottom — that is, as the father's age

(17) BELL, A. G. *The Duratiou of Life and Conditions associated with Longevity. A Study of the Hyde Genealogy.* Washington 1918, p. 57.

at death increases — and in each row of the table the average expectation of life of the offspring increases as we pass from left to right — that is, with increasing age of the mother at death. However the matter is taken, a careful selection of one's parents in respect of longevity is the most reliable form of personal life insurance.

Just how significant this is can best be realized by a comparison with the increase in expectation of life at birth which would result from the total and complete eradication of the disease tuberculosis. Dr. LOUIS I. DUBLIN (1) has lately made an interesting study of this point, with the results shown in Table 3.

TABLE III.

Average after life-time at age zero, and number of years lost because of tuberculosis. Experience of Metropolitan Life Insurance Company.

Industrial Department, 1911 to 1916. (After DUBLIN*).*

Tuberculosis	White males	White females	Colored males	Colored females
Present	46.05	51.8	37.19	38.6
Not present . .	49.53	54.4	42.14	43.7
Years lost because of tuberculosis .	3.48	2.6	4.95	5.0

It is at once apparent that even this maximum extension of life which would be expected to follow the total eradication of so important a cause of death as tuberculosis amounts to only one-fourth of the increase which follows upon having both parents live to an age of 80 or more.

Turning tho the experimental side one finds equally striking and quite different evidence that this character duration of life is definitely inherited. It has already been shown that *Drosophila* populations exhibit life curves of fundamentally the same form as

(1) « *Statistical Bulletin Metropolitan Life Insurance Co.* » Vol. 1, No. 2, p. 5, and No. 3, p. 5. 1920.

those of man, measuring in days in the one case and years in the other. The *Drosophila* life curves already presented (Figs. 2 and 3) show that there are wide differences in duration of life in different stocks of this animal. What now will happen if we cross these long-lived and short-lived forms together, and carry out a typical Mendelian experiment in respect of this character? The outcome of such an experiment has been independently described by HYDE (1) and by PEARL, (2) with entirely concordant results. What these results are is shown in Fig. 4.

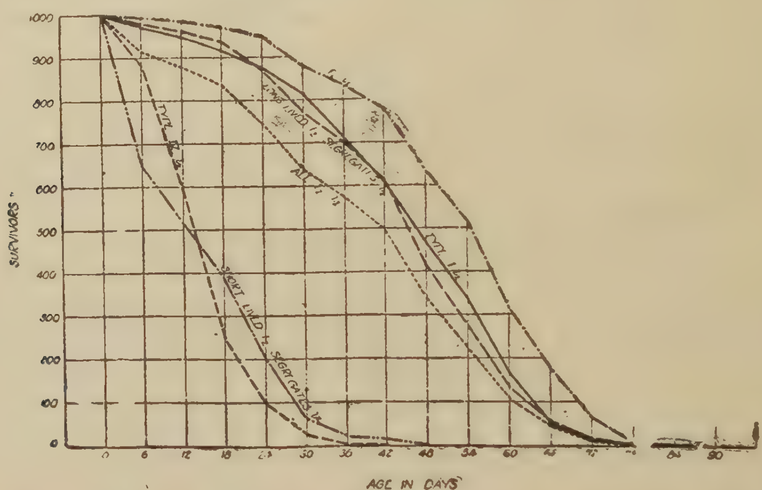


Fig. 4. Life lines showing the result of Mendelian experiments on the duration of life in *Drosophila*. Explanation in text.

In the second line from the top of the diagram, labelled «Type I l_x,» is the mortality curve for an hereditarily long-lived pure strain of individuals. At the bottom of the diagram the «Type IV l_x» line gives the mortality curve for one of the hereditarily short-lived strains. Individuals of Type I and Type IV were mated together. The result in the first offspring hybrid generation is shown by the line at the top of the diagram marked «F_1 l_x». The F_1 denotes that this is the mortaliy curve of the first filial generation from the cross. It is at once obvious that these first generation

(1) HYDE, R. R. «*Inheritance of length of life in* «Drosophila ampelophila». «Indiana Acad. Sci. Rept.» for 1913, pp. 113-123.

(2) PEARL, R. *The biology of death.* VI. *Experimental studies on the duration of life.* «Sci. Monthly». August 1921.

hybrids have a greater expectation of life at practically all ages than do either of the parent strains mated together to produce the hybrids.

The average duration of life of the Type I original parent stoch is 44.2 ± .4 days. The average duration of life of the short-lived Type IV flies is 14. 1 ± .2 days, or only about one third as great as that of the other stock. The average duration of life of the first hybrid generation shom in the F^1 l_x line is 51.5 ± .5 days. So that there is an increase in average duration of life in the first hybrid generation, over that of the long-lived parent, of approximately 7 days. In estimating the significance of this, one should remember that a day in the life of a fly corresponds, as has already been pointed out, almost exactly to a year in the life of a man.

If we consider separately the long-lived second generation group and the short-lived second generation group we get the results shown in the two lines labelled « Long-lived F_2 Segregates l_x », and « Short-lived F_2 Segregates l_x ». It will be noted that the long-lived F_2 segregates have a mortality curve which almost exactly coincides with that of the original parent Type I stock. In other words, in the second generation after the cross of the long-lived and short-lived types a group of animals appears having almost identically the same form of mortality curve as that of one of the original parents in the cross. The mean duration of life of this long-lived second generation group is 43.3 ± .4 days, while that of the original long-lived stock was 44.2 ± .4 days. The short-lived F_2 segregates shown at the bottom of the diagram give a mortality curve essentially like that of the original short-lived parent strain. The two curves wind in and about each other, the F_2 flies showing a more rapid descent in the first half of the curve and a slower descent in the latter half. In general, however, the two are very clearly of the same form. The average duration of life of these short-lived second generation segregates is 14.6 ± .6 days. This, it will be recalled, is almost identically the same average duration of life as the original parent Type IV gave, which was 14.1 ± .2 days.

Besides these Mendelian results it has further been shown by Pearl and Parker (1) that lines or strains of *Drosophila* may be

---

(1) Pearl, R. and Parker, S. L. *Experimental studies on the duration of life. II. Hereditary differences in duration of life in line-bred strains of Drosophila.* « Amer. Nat. » Vol 56. (*In press*).

isolated by appropriate breeding procedure, which differ by smaller amounts in mean duration of life or in forms of the life curve than do the stocks shown in Fig. 5. But these smaller differences in duration of life are found by repeated tests to be constant (1).

VII

By way of summary it may be said that the endeavor of this paper has been to show why and how the statistical method can be useful to medical research. That it will play an increasingly important role in this field admits of no doubt. There is no substitute for it. The time is near at hand when the investigator of medical problems whose conclusions are not grounded on an adequate quantitative base will find it difficult to get a respectful hearing in the court of scientific opinion.

In conclusion I cannot do better than to quote some remarks of the distinguished statistician, Dr. MAJOR GREENWOOD, (2) made some years ago, but still extremely pertinent. He said:

« At the present time few choose to understand the language of statistics, and many attempt to excuse themselves by the assertion that a knowledge of the language is not attainable by the average medical man. This assertion is wide of the truth. During the past few years I have had in my laboratory several workers whose preliminary training in mathematics had been neither greater nor less than that of all qualified medical men. Any of these would bear me out in the statement that a knowledge of statistical processes sufficient to enable one intelligently to adopt the modern methods requires no more application and effort than are necessary in branches of investigation universally admitted to be within the competence of all post-graduate medical workers. I do not mean to suggest that the acquirements necessary for the original investigator in pure statistical theory can be conferred upon everyone. To assert that an ordinarily intelligent man can learn to play a fair game of chess does not mean that Morphys and

---

(1) Unfortunately in this brief discussion nothing can be said of the methods used in carrying out any of the experiments with *Drosophila* the results of which are herein described, nor the care exercised to control and keep constant various factors. The reader must consult the detailed papers which have been referred to in order to get information on all points of experimental technique.

(2) GREENWOOD, M. *On methods of research available in the study of medical problems* «Lancet,» Jan. 18, 1913.

Laskers are the product of environment alone. The analogy will indeed serve us a little more. In a complex position on the chess-board it may require a master to indicate the right continuation to enable, say, black to win. But a man with an ordinary knowledge of the game will be able to follow the line of play proposed by the master and to satisfy himself of its correctness, or the reverse; he will ignore the confident assertions of those who do not even know the moves that no rational line of play can be laid down in such complex positions, that success is a matter of temperament or personal equation or « experience ».

Dott. Luigi Colomba
prof. ord. di Mineralogia, nella R. Università di Modena

La Statistica e le Scienze Naturali

In una delle sue sedute la Facoltà di Scienze della Università di Modena, su mia proposta ed a voti unanimi, deliberava di porre la Statistica fra le materie a scelta alle quali si possono iscrivere i laureandi in Scienze Naturali.

Credo utile di esporre i motivi che mi hanno indotto a fare tale proposta la cui opportunità è ora notevolmente aumentata in seguito al movimento manifestatosi nel campo della Statistica metodologica, onde ottenere che tale insegnamento venga trasferito nella Facoltà di Scienze.

È noto che sperimentalmente le leggi naturali sono ricavate mediante un criterio empirico; esse vengono cioè dedotte dall'esame di un numero più o meno grande di casi singoli, dai quali poscia, mediante processi pure più o meno ampi di integrazione, di interpolazione e di estrapolazione, si giunge a quegli enunciati che ne dovrebbero costituire le formole teoriche, considerate come pure espressioni matematiche.

In generale però i dati sperimentali, presi tanto singolarmente quanto nel loro complesso, non corrispondono mai esattamente a quelli teoricamente richiesti; essi invece, sebbene in grado molto variabile, oscillano sempre intorno ai detti valori teorici, per cui in realtà si può solo parlare di leggi limiti od approssimate. Le cause che influiscono sul loro grado di approssimazione sono molteplici e se talune, dipendendo solo da insufficienza dei mezzi di osservazione, possono considerarsi come transitorie (1), altre se ne hanno che sono direttamente colle-

(1) La insufficienza delle nostre cognizioni può anche portare ad una inesatta interpretazione di un fenomeno. Cio avverrebbe ad es. ad un individuo, avente solo cognizione del circolo e non dell'ellisse, il quale constatando che l'orbita terrestre è quasi circolare, stabilisse una legge per cui le orbite dei pianeti sarebbero circolari e considerasse quindi come deviazioni da detta legge le variazioni derivanti dall'essere tali orbite realmente ellittiche.

gate alla natura dei fenomeni che si esaminano, oppure ad alcune loro particolari condizioni. Infatti studiando lo svolgimento dei fenomeni naturali si osserva che si può in certo modo stabilire una seria continua che da quelli che sembrano obbedire a leggi di una precisione e di un rigore assoluti va ad altri i quali presentano oscillazioni talmente grandi da rendere dubbia anche la possibilità di formulare per essi una qualsiasi legge definita.

Però anche le leggi che si riferiscono a fenomeni il cui svolgimento deve considerarsi come definito in modo assoluto, si presentano a noi come semplici leggi limiti od approssimate per vari motivi come ad es. per il fatto che sperimentalmente non possiamo e forse non potremo mai operare in quelle condizioni che sarebbero necessarie per la loro completa realizzazione. Ciò appunto si verifica ad esempio nella legge di Boyle e Mariotte la quale, come è noto stabilisce che il volume di un gaz è inversamente proporzionale alla pressione a cui è sottoposto, rimanendo costante la temperatura. Ora, per quanto non vi sia alcun dubbio sulla reale esistenza di questa legge, tuttavia essa è dal lato sperimentale una semplice legge limite, perchè occorrerebbe per il suo completo avverarsi che i gaz fossero allo stato di *gaz perfetto,* cosa impossibile a verificarsi fisicamente.

Ad analoghe conclusioni si giunge considerando altre leggi fondamentali, come ad es. quella delle proporzioni definite e multiple così importante per la chimica, perchè direttamente collegata ai valori dei pesi atomici. Questa legge si enuncia dicendo che in una combinazione chimica gli elementi entrano secondo i loro pesi atomici o secondo loro multipli (il che è logico se si ammette che chimicamente l'atomo rappresenti l'ultimo grado di divisibilità dei corpi). Però anch'essa ha per noi un valore solo di legge approssimata poichè, trattandosi di dati ponderali, i valori sono sempre collegati col grado di sensibilità e di esattezza delle bilance.

In questo primo gruppo di fenomeni quindi nessuno pensa, a meno di non voler demolire addirittura le attuali basi della Scienza, che le oscillazioni e le deviazioni osservate possano infirmare in qualche modo il contenuto delle rispettive leggi, che rimangono quindi ancora assiomatiche. Così riferendoci ancora all'ultimo esempio soprariferito, quando si voglia stabilire la costituzione molecolare di una combinazione chimica partendo dalla sua composizione centesimale e non si voglia negare il

fondamento stesso della teoria atomica, i rapporti fra i vari componenti vengono sempre ricavati partendo dalla legge delle proporzioni definite e multiple e, poichè fra i dati sperimentali e quelli teorici vennero talvolta notate differenze troppo grandi per poter essere spiegate semplicemente come errori di osservazione, piuttosto che considerare falsa la detta legge, si è preferito, di applicare il principio dell' isotopismo secondo il quale i pesi atomici degli elementi chimici, possono variare, fatto questo ormai confermato sperimentalmente.

Ma oltre a queste leggi, altre se ne hanno che sono dedotte da gruppi di fenomeni i quali presentano un grado di variabilità molto maggiore nei loro dati sperimentali. E, se per talune, come ad esempio per la legge di DULONG e PETIT sui rapporti fra i pesi atomici e i calori specifici dei singoli elementi(1), si ammette ancora che si tratti di leggi aventi un valore assoluto malgrado le grandi oscillazioni che si osservano nei risultati, per altre invece occorre necessariamente di ammettere una molto maggiore ampiezza nei loro limiti di realizzazione e quindi nei loro enunciati che diverrebbero molto più larghi e comprensivi, non essendo anche da escludersi che taluna di esse possa scindersi in leggi nuove, le quali, pur derivando da un unico ceppo potrebbero in seguito divergere più o meno sensibilmente, col moltiplicarsi dei dati sperimentali che vi si riferiscono.

Ora è appunto in questi gruppi di fenomeni che i metodi statistici vengono ad assumere una grandissima importanza derivante dal fatto che ci troviamo in presenza di problemi la cui soluzione richiede una impostazione molto differente da quella richiesta per quelli precedentemente esaminati.

Infatti occorre in questi casi di ricorrere, e talvolta in condizioni molto sfavorevoli, a procedimenti di carattere puramente sintetico ed integrativo, i quali dall'esame dei singoli dati ci debbono permettere di stabilire se si tratti di fenomeni che possano venir collegati gli uni agli altri in modo da essere esprimibili mediante leggi definite, ed in caso affermativo quali

---

(1) Questa legge, che nel suo contenuto teorico si può esprimere mediante una iperbole equilatera, si enuncia dicendo che il prodotto del peso di combinazione di un elemento chimico per il suo calore specifico porta ad un valore costante, detto calore atomico ed in media uguale a 6,4. Questa costante però oscilla notevolmente.

possano essere le estensioni di dette leggi e quali i loro enunciati più logici e concordanti con quanto venne realmente osservato.

Questi risultati si possono evidentemente solo ottenere accumulando il massimo numero possibile di dati e, per un lato compiendo su di essi un lavoro accuratissimo di selezione che ci permetta di eliminare i dati meno attendibili, e per altro lato compiendo un processo di elaborazione che ci conduca a considerare le questioni dal punto di vista più largo ed obbiettivo possibile, il che disgraziatamente non succede spesso nello studio dei fenomeni naturali che vengono piuttosto studiati da un punto di vista molto ristretto, immediato e soggettivo.

I fenomeni appartenenti a questi tipi sono frequentissimi nel vasto campo delle Scienze Naturali, per cui non vi è alcuna difficoltà nella scelta degli esempi.

Io mi limiterò ad accennare a taluni presi quà e là in campi anche molto differenti, onde far vedere a quali risultati si sia giunti in alcuni casi in cui i detti metodi vennero ampiamente impiegati ed a quali, a parer mio, si potrebbe giungere in altri casi, se si evitasse di studiarli, come venne fatto fino ad ora, da un punto di vista troppo limitato ed unilaterale. Si vedrà così come si siano ricavati relativamente a problemi talvolta anche molto complessi dei risultati che, non solo hanno permesso di considerare le questioni in essi trattate con tutta la ampiezza richiesta, ma pur anche hanno servito a dimostrare come molte deviazioni ed oscillazioni dipendessero da una incompleta valutazione dei dati sperimentali, perchè non sempre erano stati sufficientemente definiti i fondamenti ed i limiti delle leggi che si consideravano.

In altri casi poi molto più difficili, perchè riferibili a fenomeni assai oscuri nelle loro cause e nel loro svolgimento, si poterono ottenere dei risultati non indifferenti che permisero di rendere meno arbitrarie le loro possibili curve di integrazione.

Vediamo ad esempio la legge dell'isomorfismo come era stata primamente definita da Mitscherlich. Secondo questa legge i corpi che hanno analogia di composizione chimica e di struttura molecolare sono capaci di dar luogo a cristalli misti i cui caratteri variano fra quelli proprii dei cristalli dei singoli corpi in rapporto alle loro rispettive proporzioni.

Questa legge quando venne emessa aveva tutta l'apparenza di una legge assoluta; ciò proveniva però solo dal fatto che

era stata fondata su un numero troppo esiguo di casi eccezionalmente favorevoli ed infatti in breve si ebbero tante deviazioni e talmente grandi che essa venne da taluno addirittura negata supponendosi che i casi ad essa favorevoli dipendessero semplicemente da fortuita coincidenza.

Le cose cambiarono però nuovamente di aspetto quando, avendosi potuto accumulare un grandissimo numero di osservazioni, queste poterono essere convenientemente elaborate, dopo di essere state sottoposte ad un diligente lavoro di epurazione. Ed invero, mentre per un lato si potè affermare la reale esistenza della legge dell'isomorfismo, fissandone in modo ben chiaro i suoi veri limiti, per altro lato si potè dimostrare che molte deviazioni dipendevano dal fatto che erano stati erroneamente riuniti dati riferentisi ad altre leggi le quali, pur avendo un fondamento comune con quella dell'isomorfismo, se ne staccano poscia nettamente.

Tali sono appunto le leggi dell'isogonismo e delle soluzioni solide; di queste la prima ammette che corpi dotati di una certa analogia di struttura molecolare e di volumi molecolari, molto prossimi, possano cristallizzare in forme molto vicine, pur non potendo dar luogo a cristalli misti; la seconda stabilisce la possibilità di analogie fra i corpi liquidi ed i corpi solidi nel loro comportamento per ciò che si riferisce al fenomeno delle soluzioni che potrebbero anche avvenire direttamente fra corpi solidi colle stesse leggi colle quali avvengono fra corpi liquidi e fra corpi liquidi e solidi.

Di queste due leggi la prima non ha per il momento che una importanza limitata, essendo piuttosto indeterminata; la seconda invece ha ormai acquistato un grande sviluppo, non solo nel campo della pura Chimica-fisica, ma pur anche in quello della Metallografia riguardo alle questioni riferentisi alla costituzione delle leghe metalliche ed in quello della Mineralogia, dove è stata, e spesso con molto successo, applicata per spiegare la costituzione di molte specie che furono considerate come soluzioni solide di corpi differenti in proporzioni variabili non potendosi ad esse applicare la legge dell'isomorfismo per la mancanza di analogie chimiche.

Un altro esempio degno di nota è quello riguardante l'aumento graduale di calore che si osserva coll'aumentare delle profondità terrestri, in rapporto colla esistenza di un calorico interno proprio del nostro globo da taluni considerato come un

residuo del calore siderale della Terra e da altri invece come dovuto a fenomeni esotermici di natura chimica e fisica che avverrebbero nell' interno del globo.

Però, siccome le nostre ricerche non hanno fino ad ora raggiunto che una massima profondità di circa 2.000 metri sopra 6.500.000 che rappresentano la lunghezza media del raggio terrestre, si comprende facilmente come, per quanto riguarda il probabile modo di svolgersi del fenomeno nelle zone molto profonde, non si possano avere risultati concreti, essendo necessario di ricorrere a pure formule di estrapolazione.

Tuttavia anche in questo caso qualche buon risultato si potè avere accumulando il massimo numero di osservazioni e sottoponendole ad un diligente lavoro di analisi. Infatti si potè, non solo stabilire che detto aumento non è arbitrario, poichè, pur avendosi nei casi singoli notevoli oscillazioni, si è constatato che l' aumento di 1° corrisponde in media ad un aumento di profondità di circa 30-40 metri (grado geotermico), ma pur anche, come appunto fece il Fritz, si è giunti alla conclusione che al centro della Terra la temperatura non può essere superiore a circa 3.500°, mentre dapprima si era giunti da taluni a cifre fantastiche superiori ai 180.000°. (1)

Porterò, togliendolo dalla Biologia, un ultimo esempio connesso colle ricerche biometriche che a mio parere conforta la mia affermazione secondo cui molte volte la mancanza di buoni risultati dipende dal modo troppo unilaterale col quale molti problemi vengono studiati, senza che si cerchi di utilizzare nel modo migliore possibile il materiale di studio che si ha a disposizione.

Poichè si tratta di un argomento già molto trattato e discusso per opera di numerosi autori come Volterra, Pearson, Weldon, Galton, Quételet, De Vries, Camerano, ecc. specialmente in quanto si riferisce alle osservazioni dei caratteri so-

---

(1) Nei primi tentativi si era infatti partiti da un' equazione lineare del tipo $KP=T$, essendo P la profondità, T la temperatura e K una costante rappresentata nel caso specifico da $\frac{1}{G}$, essendo G il grado geometrico. Posteriormente invece si partì dalla equazione di secondo grado $T^2=1{,}8(P\text{-}50)$ dalla quale, sostituendo a P il raggio terrestre e non tenendo conto del termine -50 perchè del tutto trascurabile di fronte al valore di P si ottiene, al centro della terra $T=\sqrt{1{,}8\times 6.500.000}$.

matici, mi limiterò ad accennare a quelle ricerche biometriche istituite allo scopo di vedere se sia possibile di stabilire a quale specie appartenga un individuo, mediante uno studio comparato delle varie parti del suo corpo, rimanendo però sempre nell'ambito di specie molto vicine.

Il metodo consiste nel prendere una data parte del corpo dell'individuo come unità di misura, ricavando poscia le dimensioni comparate delle altre parti. Ripetute queste misure su molti individui appartenenti alle specie considerate si costruiscono le curve di frequenza e si osserva se e come esse si differenziano le une dalle altre.

Ora le dette ricerche hanno dimostrato che nel caso di specie molto vicine le curve di frequenza, aventi il solito tipo campanulare, sono molto simili e per di più si sovrappongono quasi completamente le une alle altre, per cui i campi di saturazione delle singole specie vengono a confondersi.

Io credo che questi risultati così poco confortanti siano dipesi sopratutto dalla imperfetta impostazione del problema, la cui soluzione nei limiti prima accennati già a *priori* si poteva intravedere, essendo logico che fra specie molto vicine si abbiano minime differenze biometriche, dato che gli stessi biologi affermano che è molto difficile di stabilire in detti casi con esattezza le differenze specifiche.

Non deve quindi far stupire se le ricerche compiute in questo senso, non portino a quei risultati conclusivi che invece molto probabilmente si otterrebbero se il problema fosse studiato sotto un aspetto più generale, come ad esempio quello di ricercare quali siano i rapporti di affinità che esistono fra i caratteri specifici ed il complesso dei dati biometrici nelle varie specie che entrano a formare un genere od anche una famiglia, vale a dire fra gruppi di specie più differenziate. Invero in tale caso si potrebbe vedere se le singole curve di frequenza, certamente molto più differenziate che non nel caso di specie affini, presentino o no nel loro sviluppo qualche relazione che porti a rapporti precisi e ben definiti.

Non credo necessario di spingermi più oltre perchè suppongo che sia sufficiente quanto ho detto per dimostrare la necessità per i naturalisti, considerati nel significato più esteso della parola, di conoscere i metodi della Statistica come mezzo di indagine.

Tale convincimento deriva anche dalla impossibilità di separare nettamente i fenomeni naturali dai fenomeni sociali, non

essendovi alcun motivo perchè la serie continua che dai fenomeni riferentisi alla così detta materia bruta, va gradatamente fino a quelli proprii della materia organizzata, si arresti quando si giunge all' uomo. Invero, se anche in molti casi la attività cosciente dell' uomo ha potuto e può notevolmente influire sullo svolgimento di molti fenomeni ed anche può dar origine a fenomeni nuovi, non bisogna scordare che ve ne sono molti in cui tale sua azione cosciente ha una importanza nulla o molto limitata, e di tali fenomeni sono piene la antropologia, la antropometria ed in parte anche la demografia.

J. BOKALDERS

Lettlands Agrarproblem

Lettland vor dem Kriege war seit altersher typisch das Land des Grossgrundbesitzes, denn die 1300 privaten Güter(1) nahmen 48% des Gesamtareals, oder ungefähr $3\frac{1}{2}$ Millionen ha ein. Dem Staate gehörten vom Gesamtareal 10% oder ca 650.000 ha, wovon 80% Wälder waren. Die 225 Staatsgüter hatten durchschnittlich eine Grösse von ca 300 ha. Die 150 Pastorate bildeten 1% oder ca 68.000 ha. bei einer mittleren Grösse von ca 100 ha; 1,5% oder 89.000 ha. waren in Nutzung der Städte u. verschiedener Institutionen, und nur 39,5% = ca 2,6 Millionen ha Gesamtareal entfiel auf kleine u. mittlere bäuerliche Wirtschaften.

Eine Erläuterung für die Entstehung dieser Agrarverhältnisse geben die eigenartigen Gesetze bei Aufhebung der Leibeigenschaft, sowie die lokalen Bauerngesetze, welche von den örtlichen adligen Verwaltungsorganen - den Landtagen - ausgearbeitet u. erlassen wurden.

Erst nach Aufhebung der Leibeigenschaft - im Anfange des 19. Jahrhunderts - fing der Adel an, *auf freie Vereinbarung* Land zu verkaufen. Die Regierung bestimmte für Privatgüter aber weder die Zeit noch den Preis für den Verkauf des Bauernlandes. Nur der Verkauf der dem Staate gehörigen Bauernländereien wurde durch ein allgemeines russisches Bauerngesetz geregelt. Daher war vor dem Kriege noch ein beträchtlicher Teil des Bauernlandes unverkauft und die Anzahl der in bäuerlichem Besitz befindlichen

(1) Man unterschied in Lettland 3 grosse Gütergruppen: 1) Privatgüter, 2) Staatsgüter u. 3) Pastorate, d. h. Güter in Nutzung der landischen Pastore.

Lettland zerfällt in 3 Provinzen: I. *Livland*, bestehend aus 4 Kreisen; II. *Kurland*, welches aus 2 Teilen besteht: a) Kurland mit 5 Kreisen u. b) Semgallen ebenfalls mit 5 Kreisen u. III. *Lettgallen*, aus 3 Kreisen bestehend.

Einheiten erreichte nur 90.000. Eine mittlere Bauernwirtschaft oder ein Gesinde in den Provinzen Livland und Kurland hatte ungefähr 30 - 40 ha; in der Provinz Lettgallen (den 3 Provinzen des früheren Gouvernements Witebsk) aber erreichten die Bauernwirtschaften nur die Grösse von 7 - 8 ha. Ausser diesen Wirtschaften gab es in stark bewohnten Gegenden, in der Nähe grösserer Städte und längs dem Meeresstrande auch noch *kleinere* Wirtschaften.

Durch die historische Entwicklung des Landes, so wie auch durch den Einfluss der adligen Grossgrundbesitzer auf die agraren Verhältnisse des Landes kam der grösste Teil der Landbewohner nicht zu eigenem Lande. Der kleinere Teil derselben pachtete infolgedessen Land von den Gütern oder den grösseren Bauernwirtschaften. Der weitaus grössere Teil der Landlosen war indessen gezwungen, sich als Landarbeiter auf den Gütern oder den grösseren Bauernwirtschaften zu verdingen, in die grösseren Städte zu ziehen, oder nach Russland auszuwandern. Daher betrug die Einwohnerzahl in Lettland vor dem Kriege 40 Personen pro □ klm. Jetzt ist die Einwohnerzahl bis auf 25 Personen pro □ klm. gesunken. Die Gesamtbevölkerung Lettlands beträgt zurzeit 1,9 Millionen.

Die mittlere Grösse des privaten Grossgrundbesitzes überstieg 2.000 ha, doch gab es Güter mit bedeutend grösserem Areal, unter diesen eines - das Gut Dondangen - sogar mit 70.000 ha.

Von allen privaten Gütern in Livland waren 70 % grösser als 1000 ha, was 94 % des Gesamtareals der Privatgüter, oder beinahe die Hälfte ganz Livlands ausmachte. In Kurland waren 47,5 % dieser Güter grösser als 1000 ha d. s. = 90 % des Gesamtareals der Privatgüter, oder 37 % vom Flächenraum Kurlands. In Lettgallen nahmen die Privatgüter von je über 1000 ha Grösse 561.545 ha ein, oder ca 77 % des ganzen Privatbesitzes, was etwas weniger als die Hälfte Lettgallens ausmacht. Privatgüter von 5.000 - 10.000 ha umfassten in Livland 25,68 % des Gesamtareals dieser Güter, in Kurland 24,58 %; aber die Privatgüter in solchem Umfange in Lettgallen = 7,65 %. Diese Güter hatten den Charakter von Latifundien und umfassten 1,4 Millionen ha oder mehr als 1/5 ganz Lettlands. Güter von über 10.000 ha Grösse repräsentierten in Livland 18,25 % (208.562 ha) des Areals aller Privatgüter; der Prozentsatz für Kurland war 24,45 % (381.470 ha) und für Lettgallen sogar 34,19 % (250.102 ha).

Die grossen privaten Güter, welche von einem Zentrum aus nicht bewirtschaftet werden konnten, zerfielen meist in mehrere

Wirtschaftseinheiten, die teils vom zentralen Hof bewirtschaftet, teils verarrendiert oder auch zum Teil an Landarbeiter als Lohn vergeben wurden. Diese grossen Besitzlichkeiten waren bis zur letzten Zeit fast ausschleisslich im Besitze des Adels, welcher die landischen Verhältnisse ordnete. Von allen 414 Privatgütern Livlands waren in nichtadligen Händen nur ca 50 Güter mit ungefähr 60.000 ha, also nicht voll 1/16 des Landes. Ebenso gehörten auch in Kurland 92,2 % aller Privatgüter dem Adel, und die gleichen Verhältnisse waren in Lettgallen. Diese adligen Grossgrundbesitzer waren gewöhnlich anderer Nationalität als die eingeborenen Bewohner des Landes - die Letten: in Livland und Kurland waren es Deutsche, und in Lettgallen - Polen.

Zu der durch sozialen Antagonismus in der Agrarfrage hervorgerufenen Unzufriedenheit gesellten sich auch nationale Gegensätze, welche sich durch die den Grossgrundbesitzern zustehenden Privilegien noch verschärften. Diese feudalen Privilegien, welche seit dem Mittelalter bis zum Weltkriege bestanden, waren: das Vorrecht des Bierbrauens und des Spiritusbrennens, das Jagd — und Fischereirecht. In der Verwaltung des Landes und in den Selbstverwaltungsorganen (Landtagen) standen den Grossgrundbesitzern in Gemeinsamkeit mit russischen Regierungsbeamten ebenfalls das Bestimmungsrecht und das ausschlaggebende Wort zu. (1)

Diese Agrarverhältnisse und die sozialen Zustände riefen schon vor dem Weltkriege unter der Landbevölkerung Unzufriedenheit mit den Verwaltungszuständen und der bestehenden Agrarordnung hervor, was besonders deutlich in den Unruhen i. J. 1905 zutagetrat.

Durch den Weltkrieg, die russische Revolution und endlich durch die Gewaltherrschaft der Bolschewisten wurde der Grossgrundbesitz sehr in Mitleidenschaft gezogen. Viele Besitzer hatten ihre Güter verlassen, andere verliessen sie infolge der Kämpfe bei Riga und Wenden und wegen der Bermondtaffäre im Jahre 1919.

Darum hatte die Regierung Lettlands noch vor dem Inkrafttreten des neuen Agrargesetzes schon fast die Hälfte aller Güter Lettlands unter ihre Aufsicht genommen und zwar: die von ihren Besitzern und deren Bevollmächtigten verlassenen und die infolge der Bermondtaffäre sequestrierten Güter. Die Anzahl solcher von der Re-

---

(1) Solche Güter, welche das nach dem Gesetz bestimmte Minimum (300 ha. land irtschaftlich genutzer Fläche) aufwiesen und deren Besitzer bestimmte oben erwähnte Vorrechte genossen und sämtlich Glieder der provinzialen Landtage waren, hiessen Rittergüter.

gierung übernommener Güter betrug zu Ende des Jahres 1920 - 629. Güter und *16 Pastorate.*

Das Areal aller dieser Güter betrug ca 800442 ha. 133 dieser Güter mit 182.671 ha Gesamtareal, darunter 22.213 ha Acker, bewirtschaftete die Regierung selbst, während sie die übrigen auf kurze Zeit an Gross = oder Kleinpächter abgab.

Gegen Ende des Jahres 1919 begannen schon die Vorarbeiten für die Landaufteilung, die dann während des ganzen Jahres 1920 fortgesetzt wurden; die abgeteilten Grundstücke wurden verpachtet und die Vorarbeiten für die in Aussicht genommene Agrarreform in Angriff genommen. Unter solchen Umständen entstand der erste Teil des jetzt bestehenden Agrarreformgesetzes.

Dieser erste Teil «über den staatlichen Landfonds» ist am 16. September 1920 von der Konstituante angenommen worden, der II. Teil «über die Nutzung des staatlichen Landfonds» — am 21. Dezember 1920; der IV. Teil « über die Landeinrichtungskomitees » — am 17. September 1920, und der III. Teil «enthaltend die Bestätigung der Agrarordnung» am 3. Mai 1922. Ausser den oben erwähnten Teilen des Agrargesetzes gibt es noch eine ganze Reihe von verschiedenen Instruktionen, die vom Landeinrichtungskomitee, das über die Nutzung des staatlichen Landfonds die Aufsicht hat und die Landverteilungs = und Zuteilungsarbeiten leitet, erlassen worden sind, um auf administrativem Wege die Agrarreformarbeiten in der Praxis durchzuführen un zu regulieren.

Das Zentrale Landeinrichtungskomitee besteht aus 13 Gliedern, von denen die eine Hälfte von der Konstituante ernannt u. die andere Hälfte-mit dem Minister der Landwirtschaft als dem Vorsitzenden des Komitees-von der Regierung eingesetzt wird.

Die Ziele der Agrarreform sind laut Gesetz folgende:

1) Die Einrichtung neuer Wirtschaften und die Erweiterung schon bestehender Kleinwirtschaften,

und 2) die Befriedigung verschiedener wirtschafticher, sozialer und kultureller Bedürfnisse und Erweiterung der Städte u. Flecken.

Der I. Teil des Agrarreformgesetzes bezieht sich direkt auf die Aufteilung des bisherigen Grossgrundbesitzes und die Gründung neuer Wirtschaften, während der II. Teil ohne direkte Angaben sich weitere soziale und kulturelle Ziele steckt, die als Folgeerscheinungen einer weitgehenden Agrarreform zu betrachten sind.

Um die oben erwähnten Ziele durch das Agrarreformgesetz zu erreichen, ist speziell der *staatliche Landfonds* gegründet.

Dieser Landfonds besteht:

1) aus dem Staatsbesitz (Güter u. Wälder),

2) aus dem Privatbesitz (Privatgüter mit Ausnahme der von ihnen abgeteilten u. verkauften Bauernhöfe.)

u. 3) aus den Pastoratsländereien.

Die unverkauften Bauernhöfe, so wie die in der deutschen Okkupationszeit zu Kolonisationszwecken zugewiesenen und abgeteilten Ländereien, der Besitz der russischen Bauernagrarbanken mit den von ihnen gekauften Bauernländereien, sind vollkommen dem Landfonds einverleibt. Nicht enteignet wird der den früheren Landbesitzern nicht zu enteignende Teil, der nicht mit dem Gutszentrum gleichbedeutend ist, einer *mittleren* verkauften Bauernwirtschaft gleichkommt und in jedem einzelnen Falle durch die Regierung resp. durch das Landwirtschaftsministerium zugewiesen wird. Die Grösse eines solchen *nicht zu enteignenden* Teiles ist auf 50-100 ha bestimmt. Ferner sind vorläufig noch nicht enteignet: die Güter u. Ländereien, welche den Städten, Flecken, Kreisen und Dörfern gehören. Zugleich kann mit dem Lande, gegen Entgelt, auch das der Regierung nötige u. taugliche Inventar enteignet werden.

So sind durch das neue Agrargesetz dem staatlichen Landfonds fast $^2/_3$ allen Landbesitzes Lettlands zugezählt worden, d. i.:

landw. genutzt. Fläche:	1.654.739 ha,
Wald . . . . :	1.496.884 ha,
Unland . . . . . . :	558.791 ha,
Gesamtareal:	3.710.414 ha.

In diese Summe sind *nicht* hineingerechnet die durch Grenzregulierungen und Friedensverträge Lettland zugefallenen Gebiete. Wenn man von obigem Areal die den früheren Besitzern und Pastoren belassenen Ländereien abrechnet, so verbleiben dem Landfonds: 3.573.000 ha. Den grössten Teil des Landfonds (81.3 %) bilden die ehemaligen Privatgüter mit:

1.409.501	ha landw. genutzt. Fläche,
1.128.446	ha Wald,
477.902	ha Unland
3.015.849	ha Gesamtareal;

den nächst kleineren Teil (16,9 %) die *Staatsgüter* und *Wälder* mit:

188.782	ha landw. genutzt. Fläche,
362.374	ha Wald
76.578	ha Unland
627.734	ha Gesamtareal;

und den kleinsten Teil (1,8 %) - die früheren *Pastorats- u. Kirchenländereien* mit:

56.456	ha landw. genutzt. Fläche,
6.063	ha Wald,
4.311	ha Unland
66.830	ha Gesamtareal.

Dieser oben erwähnte Landfonds ist vom Staate folgendermassen zu nutzen:

1) Die Wälder, Gewässer, das Unland, historische Gegenden, Ländereien mit Naturschönheiten u. Bodenreichtümern, oder solche, welche archäologischen Wert haben, bleiben als Staatseigentum in staatlicher Verwaltung und zwar in einem Umfange, der durch ein besonderes Gesetz, das aber noch nicht herausgegeben ist, festgesetzt werden wird. Eben beträgt der Staatsbesitz — an Wäldern u. Unland — ca 2 Millionen ha. Jeder neufundierten Wirtschaftseinheit können bis 5 ha. Unland u. Wald zugeteilt werden, wobei der Wald 3 ha betragen kann. Gleichfalls enthalten auch die den früheren Besitzern zugeteilten Landstücke Wald u. Unland. Der Staatsbesitz hat sich durch das neue Agrargesetz *verdreifacht* und zwar ist er von 630.000 ha auf 2 Millionen ha angewachsen.

2) Die landwirtschaftlich zu nutzenden Ländereien des staatlichen Landfonds sollen in erster Reihe zur Gründung neuer Wirtschaften verwandt werden. Für diese Wirtschaften hat das Agrargesetz als Maximum 27 ha bestimmt, davon können 22 ha landwirtschaftlich genutze Ländereien sein (Acker, Wiese, Weide) und die übrigen 5 ha dürfen Wald u. Unland sein.

In den staatlichen Landfonds sind vorläufig auch alle langbefristeten Pachtgrundstücke aufgenommen worden. Diese sind aber nicht aufzuteilen, sondern nach Regulierung der Grenzen endgültig den bisherigen langjährigen (wenigstens 25 Jahre) Pächtern zuzu-

teilen. Diese Pachtgrundstücke unterscheiden sich wirtschaftlich durchaus nicht von den verkauften Bauernwirtschaften (Gesinden) und haben auch einen genügenden Prozentsatz (12 %) an Wald u. Unland. In Livland, Kurland und Lettgallen sind die auf Hofesland gelegenen Bauernwirtschaften nicht zu teilen, wenn sie den Charakter von Gesinden tragen und die Grösse von 100 ha nicht übersteigen.

Schon bestehende Kleinwirtschaften, deren Areal das der neufundierten Wirtschaften nicht erreicht, können vom Landfonds bis auf 27 ha Gesamtareal vergrössert werden. Kleinsvirthschaften sind laut Agrargesetz, Wirtschaften mit bis 15 ha Gesamtareal. Schon vor dem Kriege gab es in Lettland viele solcher Kleinwirtschatten, besonders in Lettgallen.

Das Areal, welches zur Gründung neuer Wirtschaften und zur Vergrösserung schon bestehender zur Verfügung steht, beträgt ca 1.655.000 ha landwirtschaftlich zu nutzender Fläche, — ohne Wald u. Unland. Da nun für jede Wirtschaftseinheit 5 ha an Wald und Unland hinzuzurechnen sind, was für ca 60.000 Wirtschaften 300.000 ha ausmacht, so stehen also dem staatlichen Landfonds in nächster Zukunft ca 2.000.000 ha für die Zwecke der Agrarreform zur Verfügung. Der übrige Teil des Landfonds bleibt Staatseigentum u. wird erst allmählich für die Zwecke der Agrarreform ausgenutzt werden können und zwar nach Trockenlegung von Sümpfen und nach Ausführung von umfassenden Bodenmeliorationsarbeiten.

Nach dem Agrarreformgesetz haben ein Anrecht auf Land aus dem staatlichen Landfonds alle Bürger Lettlands zwischen 18 und 65 Jahren, welche *kein* Land oder *weniger* als 22 ha besitzen u. welche sich verpflichten, dass ihnen zuzuteilende Land zu bewirtschaften. Andere Bedingungen stellt der Staat den Lanfordernden nicht.

In den Jahren 1919 u. 1920, als die Regierung zum ersten Mal die Landaspiranten registrierte, meldeten sich unter dem Einflusse der ungeklärten politischen Lage nur 43.000 Personen (mit den Familiengliedern 72.474) in Livland u. Kurland zusammen. (Lettgallen und auch ein Teil Kurlands waren damals noch von Russland okkupiert). Diese Registrierung ergab folgende Daten:

2/3 aller Landaspiranten hatten das zur Einrichtung der Wirtschaft nötige Inventar und zwar: 61,07 % hatten ein oder mehrere Pferde, 81 % — Kühe, 66 % — Pflüge und 67 % — Wagen, da 55 % aller Landwünschenden bisherige Pächter oder Halbkörner

waren. Im Jahre 1921 ist die Zahl aller Landaspiranten bis auf 100.000 Personen angewachsen und in manchen Gegenden macht sich ein relativer Landmangel fühlbar. Während in Livland u. Kurland die Landaspiranten zum grössten Teil Landlose sind, sind es in Lettgallen meistenteils Besitzer von Kleinwirtschaften. So waren z. B. im Ludsenschen Kreise, nur 14,84 % aller Landaspiranten Landlose, während 68,05 % — Personen waren, welche Land in Pacht oder zu eigen besassen. Im Durchschnitt betrug die Landforderung im Ludsenschen Kreise 4 ha pro Person. Die Pächter hatten grösstenteils bis 15 $^1/_3$ ha Land von den Gütern oder den Gemeinden in Pacht, wünschten aber fast alle bis 20 ha pro Wirtschaft. Nur etwas mehr als 1/4 (28, 8 %) aller Landaspiranten des Ludsenschen Kreises hatten *kein* Pferd; 62,68 % hatten zu *einem* Pferde, die übrigen 8,58 % hatten *mehr* Pferde; 8,40 % der Landwünschenden hatten kein Vieh; 20,37 % — keine Pflüge und 23,32 % — keine Wagen. Es ist somit die Hauptaufgabe der Agrarreform in Lettgallen, die Kleinwirtschaften zu vergrössern und Einzelhöfe einzurichten. Ferner wäre die geringe Zahl der Landlosen mit Land zu versorgen, wenn auch durch Verpflanzung in schwächer bewohnte Gebiete Livlands u. Kurlands.

Das Agrarreformgesetz erkennt allerdings jedem Bürger Lettlands die Berechtigung auf Land zu, doch wird man in der nächsten Zukunft nur einem Teile — wenn auch dem grössten — der Aspiranten gerecht werden können.

Darum teilt das Gesetz alle Landaspiranten in Kategorien ein. Zunächst ist vorgesehen, die Ansprüche der Kleinwirtschaften zu befriedigen, wobei die lettländischen Freiheitskämpfer den Vorzug haben.

Zur I. Kategorie gehören vor allem die Landaspiranten der *betreffenden* Gemeinde: die Kavaliere des Bärentöterordens (1), die Familienglieder der in der lettländischen Armee Gefallenen und die Kriegsinvaliden.

Zur II. Kategorie gehören die Landwünschenden aus anderen Gemeinden: Krieger, welche wenigstens 1/2 Jahr in der lettländischen Armee oder einem lettischen Schützenregimente gedient und an den Befreiungskämpfen teilgenommen haben, die Angehörigen der Gefallenen und die Invaliden.

---

(1) Die Kavaliere des Bärentöterordens sind Krieger, welche an den lettländischen Befreiungskämpfen teilgenommen haben und dafür mit dem im Jahre 1920 gestifteten höchsten Kriegsorden Lettlands ausgezeichnet worden sind.

Zur III. Kategorie gehören die übrigen Landlosen der örtlichen Gemeinde,

zur IV. Kategorie — die Landlosen aus anderen Gemeinden, die das *nötige* Inventar haben

u. zur V. Kategorie — alle übrigen Landwünschenden *ohne* Inventar.

Vor den oben angeführten, in Kategorien eingeteilten Land - wünschenden muss jedoch Land zugeteilt werden:

1) den Verwaltungs, Selbstverwaltungs u. öffentlichen Institutionen, so wie für soziale, u. kulturelle Zwecke;

2) zur Erweiterung der Kleinwirtschaften und zur Gründung bisher angeforderter neuer Wirtschaften;

3) den langjährigen Pachtstellen, welche den Charakter von Gesinden haben und nicht weniger als 25 Jahre von örtlichen Landwirten bewirtschaftet worden sind;

u. 4) zur Abrundung der nicht zu enteignenden Wirtschaften und zur Liquidierung der Streustücke.

Diesen ausserhalb der obigen 5 Kategorien stehenden Gruppen sind im Jahre 1920 26 % aller verteilten Landstücke zugewiesen worden. Auf die I. Kategorie entfallen: 39 Einheiten (1.57 %) der verteilten Landstücke; auf die II. Kategorie: 501 Einh. (20,12 %); auf die III. Kategorie: 1.104 Einh. (44.34 %); auf die IV. Kategorie: 175 Einh. (7.03 %); und auf die V. Kategorie: 11 Einh. (0.44 %).

Bis zum 1. Januar 1922 sind in Liv- u. Kurland 1020 Grundstücke mit 787.948 ha Gesamtareal eingeteilt u. ihre Teilungsprojekte bestätigt worden. Von diesem Gesamtareal ist der grössere Teil — 385136 ha. — für neufundierte Wirtschaften verwandt; 42958 ha. sind den langjährigen Pächtern belassen. 46468 ha sind für Selbstverwaltungs — u. soziale Bedürfnisse verwandt und nur 6307 ha sind den Kleinwirtschaften zugeteilt worden: die übrigen, mehr als 300.000 ha, sind dem Landfonds verblieben.

Bis zum Jahre 1922 sind 22966 Wirtschaften von durchschnittlich 15-16 ha Grösse neu eingerichtet worden, 4.46 % (1024 Wirtschaften) sind zu 2 ha gross; 3.83 % (879 Einh.) — zu 2-5 ha gross; 3.73 % (857 Einh.) — zu 5 - 10 ha gross; 14.61 % (3355 Einh.) — zu 10-15 ha gross; 60.91 % (13989 Einh.) — zu 15-22 ha gross; u. 12.46 % (2869 neue Wirtschaftseinheiten) grösser als 22 ha. Fast 3/4 aller neu gegründeten Wirtschaftseinheiten sind grösser als 15 ha.

In derselben Zeit sind in Lettgallen ca 80 Dörfer mit 18236 ha Gesamtfäche in Einzelhöfe aufgeteilt u. ca 1700 einzelne Wirtschaften gegründet worden.

Im Laufe von 2 Jahren sind ca 30.000 Wirtschaften — teils *neu* gegründet, teils bestehende erweitert u. abgerundet — und teils neu vermessen worden. Von diesen sind bis zum I. Januar 1922 — 10.000 Wirtschaften den Eigentümern schon zu eigen überwiesen worden und zwar sind darunter 8.000 *neu fundierte* Wirtschaften und 2.000 frühere Pachtgesinde. Nach dem Agrargesetz können die vom Landfonds fundierten Wirtschaften als Privateigentum erworben oder arrendiert werden. Ein sehr grosser Teil der Neuwirtschaften wird zu eigen angefordert.

Auf gesetzgeberischem Wege ist die Frage betr. der Ordnung des Auskaufes der zugeteilten Grundstücke noch nicht gelöst. Das Agrarreformgesetz bestimmt wohl, dass das Land gegen Entgelt « *zu eigen* » abgegeben werden soll, aber die Zahlungsbedingungen werden erst noch durch ein Gesetz bestimmt werden. Dieses Gesetz nimmt als Ausgangspunkt die zur Zeit der Begründung des Landfonds (i. J. 1920) bestehenden Landpreise an. Ungeklärt ist ebenfalls noch die Frage über die Entschädigung der früheren Besitzer u. die Finanzierung der neuen Wirtschaften, über die Verpachtung und die Korraboration, d. i. die Eintragung des Objekts in die Grundbücher zur Sicherstellung des Eigentumsrechts.

Hand in Hand mit der Landaufteilung geht auch die Zuteilung der nicht zu enteignenden Teile des Grossgrundbesitzes (50-100 ha) an die früheren Besitzer. Im I. Halbhjar 1921 sind 166 Besitzern ihre *nicht* zu *enteignenden* Anteile überwieson worden und zwar: **131** Personen (97 %) das Zentrum oder Teile desselben.

Die Agrarreform hat während der Kampfeszeit begonnen und die infolge der Kämpfe u. der Gewaltherrschaft des Bolschewismus ungünstigen Verhältnisse haben einen grossen Einfluss auf ihre Entwicklung gehabt u. sie naturgemäss sehr behindert. Die Arbeiten der Agrarreform schreiten erst jetzt schneller vorwärts, aber die endgültigen Resultate derselben wird erst die Zukunft zeitigen.

Marcello Boldrini

La décroissance sénile chez l' homme et chez la femme

1. — Pour étudier les caractères quantitatifs de l' homme aux différents âges, les anthropologistes distinguent le plus souvent dans la vie extra-utérine deux périodes: la période de croissance qui commence à la naissance, et dont la durée est variable, et la période de l'âge adulte, pendant laquelle l'équilibre morphologique atteint est considéré comme pratiquement stable (1).

Cette distinction très grossière contraste avec une autre, plus subtile, adoptée par les physiologistes. Ils considèrent, dans la vie extra-utérine (la seule qui nous intéresse en ce moment) une première période, pendant laquelle les phénomènes de construction, ou anaplastiques, ont la prédominance et au cours de laquelle les fonctions se développent et se compliquent; une deuxième période d' équilibre énergétique, caractérisée par la stabilité des fonctions; une troisième période de déchéance, ou cataplastique, pendant laquelle les fonctions subissent une lente involution et où l' individu entre, peu à peu, dans cette phase de vieillesse physiologique qui aboutira à la mort naturelle.

On admet que les phénomènes de la période anaplastique, de la période d' équilibre et de la période cataplastique se

---

(1) Toute vie humaine - écrit le Dr. Apert dans un bon livre qui a paru récemment - se partage naturellement en deux grandes périodes: la période de croissance, qui englobe la vie phoetale, l' enfance et l'adolescence; la période ultérieure ou âge adulte, qui s'étend de la terminaison de la croissance à l'extrême vieillesse. Voir: *La croissance*, Paris, 1921, Introduction. Toutefois, nous devons rappeler que, bien qu' on ne l'aie pas soumise à une étude détaillée, la régression sénile a été signalée par Quetelet et par d'autres savants qui se sont occupés des variations physiques du corps humain en fonction de l'âge. Nous citerons, particulièrement S. Weissenberg. *Das Wachstum des Menschen nach Alter, Geschlecht und Rasse*, Stuttgart, 1911.

rattachent directement à la constitution somatique et réagissent, par conséquent, sur les caractères morphologiques.

Pour cela, il est indiqué d'approfondir l'analyse anthropologique de la phase d'involution, qui ne devrait pas offrir un intérêt moindre que la phase de croissance, puisqu'elle procède à l'inverse et détruit ce qui avait été créé précédemment. En outre, étant donné que l'âge adulte est le point d'arrivée de la croissance et le point de départ des phénomènes de la sénilité, c'est sur les périodes extrêmes qu'on devrait surtout concentrer son attention.

Malheureusement, les difficultés qui s'opposent à une étude quantitative systématique des phénomènes de la croissance aussi bien que de ceux de la sénilité sont tellement nombreuses, qu'il est aisé de se rendre compte pourquoi, cinquante ans après la publication de l'*Anthropométrie* de QUETELET, nous ne sommes pas encore arrivés à des conclusions définitives.

2. — Nous nous proposons d'examiner l'influence du vieillissement sur la taille et sur le poids du corps et de ses organes chez l'homme et chez la femme.

Il est aisé de comprendre que, à l'exception des dimensions externes et du poids de l'ensemble de l'organisme humain, il est matériellement impossible de suivre chez les individus au fur et à mesure qu'ils vieillissent, les modifications des caractères quantitatifs que nous voulons considérer.

D'autre part, il est très difficile de suivre la taille et le poids d'un individu depuis l'achèvement de sa croissance jusqu'à l'âge extrême; même en ne comptant pas sur l'éventualité de la mort, il est fort douteux qu'on puisse réussir à suivre un groupe d'hommes ou de femmes qui vieillissent suffisamment nombreux pour une étude statistique.

C'est pour cela que les observations directes ont toujours été extrêmement rares quoique suffisantes pour confirmer l'observation commune que la taille de l'homme qui vieillit subit une régression indubitable (1).

---

(1) Dans un livre bien connu M. VENTURI a pu indiquer la taille d'un certain nombre d'aliénés adultes des deux sexes à deux époques différentes pendant leur maladie. Ayant soumis à l'élaboration statistique ces données, dès que ce mémoir avait été livré à l'imprimerie, j'ai pu constater que les aliénés hommes, aussi bien que les aliénés femmes, accusent une dé-

Dans ce domaine pourtant comme dans celui de la mortalité, on a dû et on doit recourir, d'ordinaire, à l'observation indirecte. On fait porter ses observations sur des individus d'une population, appartenant aux différents âges et, en supposant que les valeurs moyennes des caractères quantitatifs des différents groupes observés aient été les mêmes au même âge — fin de la croissance dans notre cas — on cherche à déterminer l'influence de la variation de l'âge sur ces valeurs moyennes. Que l'influence de cet artifice puisse être considérée comme négligeable ou non, c'est un fait que nous discuterons à la fin de nos recherches.

C'est à Powys et Pearson que revient le mérite d'avoir observé les premiers, selon cette hypothèse, que la taille, après avoir atteint son maximum, commence à fléchir d'une allure rectiligne. Ils virent que ce fléchissement commence pour l'homme à l'âge de 27 ans et pour la femme à l'âge de 25 ans et que son intensité est de 34'' — 35'' tous les dix ans. Ces faits ont été confirmés ensuite par les études de Goring et de Harris, qui, comme les deux auteurs précédents, ont pu se servir d'un matériel statistique suffisamment sûr et abondant (1).

Si nous étendons ces recherches à d'autres caractères, il nous sera facile de constater que leur décroissance — toujours vérifiée par l'observation indirecte — est aussi, en général, grossièrement rectiligne, comme la décroissance de la taille, et peut être décrite par un parabole du premier ordre (droite) représentée par la formule:

$$y = A + Bx$$

dans laquelle y est la mesure du caractère considéré, exprimée en centimètres, kilogrammes ou grammes, x est l'âge exprimé en ans, A et B sont des constantes, que nous déterminons sur la base des observations recueillies, par la méthode des moindres carrés.

---

chéance annuelle d'autant plus intense que plus grave est la forme de folie par laquelle ils ont été atteint. Si les données de M. Venturi semblent donc confirmer le fait de la régression sénile, tel qu'il nous apparaît par l'observation indirecte, elles nous revèlent surtout la possibilité que l'action de l'âge soit accélerée et intensifiée par l'action d'un processus morbide. Voir: S. Venturi, *Le degenerazioni psico-sessuali*. Torino, 1892.

(1) A. O. Powys, *Data for the problem of evolution in man.* «Biometrika» 1901, pp. 30-49. K. Pearson a collaboré à la rédaction de quelques parties de ce mémoir. C. Goring, *The english Convict*. London, 1913, pp. 191-193. J. A. Harris, *Decrease in stature. Note on the medico-actuarial investigation.* «Quarterly Publication of the American Statistical Association», 1920, pp. 219-221.

Comme l'allure croissante ou décroissante de la droite dépend du signe de B, le signe de B sera négatif là où les caractères sont en décroissance.

En outre, comme l'inclinaison de la droite dépend de la valeur absolue de B, la droite accusera une chute rapide là où la valeur absolue de B négatif sera assez grande, tandis qu'elle accusera une chute lente là où la valeur absolue de B négatif sera petite.

Nous devons faire remarquer clairement que la représentation de l'allure de la décroissance par une parabole du premier ordre a une signification tout à fait empirique: elle ne prétend pas exprimer une loi quelconque, mais décrire l'involution sènile telle qu'elle apparaît comme le résultat des causes dont elle dépend et faciliter la comparaison entre les différents caractères du même sexe, ou bien entre les deux sexes pour le même caractére.

C'est surtout le premier de ces buts qui a été envisagé par Powys et Pearson lorsque ils ont fait remarquer que, dans l'étude de l'hérédité d'un caractère — par exemple de la taille — on doit comparer la taille du père et celle du fils, seulement après avoir éliminé l'influence que la différence d'âge exerce sur les mesures (1).

3. — Nous nous bornerons, par contre, au deuxième objectif dont nous avons parlé, en étudiant l'intensité de la décroissance de la taille et du poids du corps et de ses organes chez les deux sexes, toujours d'après l'hypothèse que nous avons exposée (2).

---

(1) Etant donné l'équation de la taille des hommes d'après les relevés de Boyd (voir le tableau): $y = 170{\cdot}7967 - {\cdot}0380\,x$, si la taille d'un individu âgé de 76 ans est de cm. 162, sa taille probable à l'âge de 25 ans (lorsqu'elle avait atteint vraisemblablement son maximum) sera: $T = 162 + {\cdot}0380\ (76 - 25) =$ cm. 163,94. Pour comparer la taille de cet homme avec celle d'un de ses fils âgé de 25 ans, il faut tenir compte d'une involution probable de 2 cm. environ.

(2) Malheureusement les données statistiques dont nous disposons sont très défectueuses. Elles ont été relevées, pour la plupart, sur la table anatomique et ne peuvent donc pas donner une idée exacte des dimensions et des poids tels qu'ils seraient s'il était possible de les observer sur des individus normaux. En effet, il suffit d'analyser les relevés d'autopsie pour se convaincre des altérations quantitatives profondes déterminées par la cause de decés, altérations qui affectent très différemment les divers organes et les différents tissus. En outre, étant donné la différence de probabilité que les individus des deux sexes soient frappés par les mêmes causes de decès, celles-ci

Le tableau suivant donne les équations de décroissance d' un certain nombre de caractères à partir de 25 ans environ. Les sources des renseignements statistiques sont indiquées dans la note (1).

Dans la première colonne du tableau, nous avons indiqué les différents caractères auxquels les équations se rapportent; dans les colonnes 2 et 4 les valeurs de y pour les hommes et pour les femmes; dans les colonnes 3 et 5 les rapports entre les écarts carrés moyens auxquels l'adjustement mathématique des valeurs observées donne lieu $\left(\sqrt{\frac{\Sigma d^2}{n}}\right)$ et les moyennes arithmétiques de chaque dimension ou poids (A); dans la colonne 6 enfin, le signe ♂ lorsque l' intensité de la décroissance (établie d'après le signe et la valeur absolue de B dans l' équation) est supérieure chez l'homne, et le signe ♀ lorsqu'elle est supérieure chez la femme.

4. — Une remarque générale doit être faite à propos de l' approximation obtenue par l' ajustement mathématique des séries observées.

Les écarts carrés moyens entre les intensités moyennes pour chaque groupe d'âge des différents caractères, données par l'observation, et les intensités calculées d'après l' hypothèse d' une dé-

---

exercent une influence inégale sur les caractères moyens des hommes et des femmes. (Voir: M. Boldrini. *Sul peso relativo del corpo e del cervello secondo la causa della morte.* «Rendiconti del Reale Istituto Lombardo di Scienze e Lettere», 1921. M. Boldrini. *Differenze sessuali nei pesi del corpo e degli organi umani.* «Rendiconti della Reale Accademia Nazionale dei Lincei, Classe di Scienze fisiche, matematiche e naturali», 1920).

On doit ajouter que la plupart des cadavres portés sur la table anatomique est fournie par les plus basses classes sociales. Ce fait non plus ne manque pas d'exercer une influence sur les dimensions et sur les poids. (Voir: M. Boldrini. *I cadaveri degli sconosciuti. Ricerche demografiche e antropologiche sul materiale della Morgue di Roma.* «La Scuola Positiva», 1920). Si nous ajoutons que l' état de conservation des cadavres observés peut être très variable; que la technique employée dans l' ablation des organes n' est pas la même chez tous les anatomistes; qu'il est difficile de faire des déterminations quantitatives suffisamment exactes et que la diligence des opérateurs n' est souvent que relative, nous devons nous convaincre de la nécessité d' employer avec beaucoup de prudence les données statistiques que nous offre la littérature.

(1) R. Boyd. *Tables of the weights of the human Body and Internal Organs, etc.* «Philosophical Transactions of the Royal Society», London, vol. 151, 1ère partie, 1861, pp. 241-262. H. L. W. v. Bischoff. *Das Hirngewicht des Menschen*, Bonn, 1880. A. O. Powys, J. A. Harris, ouvrages cités.

... de décroissance de quelques caractères chez l'homme et chez la femme.

Auteurs	Caractères (1)	Hommes — Valeurs de y (2)	$\left(\frac{\Sigma d^2}{nA^2}\right)^{\frac{1}{2}}$ (3)	Femmes — Valeurs de y (4)	$\left(\frac{\Sigma d^2}{nA^2}\right)^{\frac{1}{2}}$ (5)	Sexe chez lequel la décroissance est la plus intense (6)
Powys et Pearson	— Taille . . . . cm.	173.5836 — .0856 x	.0020	161.6128 — .0949 x	.0041	♀
Harris	— Taille . . . . »	174.3202 — .0076 x	—	164 9476 — .0500 x	—	♀
Boyd	— Taille . . . . »	170.7967 — .0380 x	.0048	161.0617 — .0808 x	.0059	♀
Bischoff	— Taille . . . . »	168.2823 — .0579 x	.0134	156.2479 — .0518 x	.0065	♂
»	— Poids du corps. . . . . . Kg.	47.3133 + .0429 x	.0250	43 9648 — .0254 x	.0569	♀
Boyd	— Poids du corps . . . . . »	42.4977 + .0680 x	.0311	38.8754 — .0220 x	.0172	♀
Bischoff	— Poids de l'encéphale . . gr.	1355.3500 — .4070 x	.0190	1235.6000 — .5720 x	.0234	♀
Boyd	— Poids de l'encéphale . . »	1397.6927 — 1.3052 x	.0082	1309.5995 — 2.0797 x	.0219	♀
»	— Poids du cerveau. . . . »	1214.8129 — 1.1552 x	.0078	1132 6277 + 1.6403 x	.0131	♀
»	— Poids de l'hémisphère droite »	589.7433 — .3776 x	.0286	557.9323 — .7116 x	.0171	♀
»	— Poids de l'hémisphère gauche »	600.4424 — .5103 x	.0296	560 4781 — .6651 x	.0161	♀
»	— Poids du cervelet . . . »	240.2620 — .1729 x	.0132	221.2395 — .1610 x	.0098	♂
»	— Poids du pont et de la moelle »	29.7372 — .0252 x	.0322	28.1340 — .4763 x	.0273	♀
»	— Poids de la moelle épinière »	31.6154 — .0181 x	.0434	31.7925 — .0677 x	.0422	♀
»	— Poids du poumon droit . . »	911.7908 — 2.3672 x	.1100	725.6735 — 3.4212 x	.0968	♀
»	— Poids du poumon gauche. »	834 6206 — 2.8145 x	.0893	536.6560 — 1.3041 x	.0248	♂
»	— Poids du cœur . . . . . »	250.2552 + 1.4387 x	.0414	226.8314 + .8292 x	.0295	♀
»	— Poids de l'estomac . . . »	168.2274 — .1619 x	.0510	153.7932 — .0902 x	.0720	♂
»	— Poids du foie . . . . . »	1875.6029 — 7.4843 x	.0468	1662.6982 — 7.5410 x	.0208	♀
»	— Poids de la rate . . . . »	219.6321 — 1.2151 x	.0613	207.3142 — 1.3880 x	.0428	♀
»	— Poids du pancréas . . . »	111.4957 — .3056 x	0584	93.0246 — .2430 x	.0432	♂
»	— Poids des deux reins . . »	342.1813 — 1.1289 x	.0644	313.8346 — 1.3506 x	.0222	♀
»	— Poids du rein droit . . . »	162 3420 — .4930 x	.0462	145.8582 — .5386 x	.0362	♀
»	— Poids du rein gauche . . »	167.4718 — .3595 x	.0455	153.2234 — .6055 x	.0291	♀
»	— Poids des capsules surrénales »	20.4556 — .0060 x	.0896	18.7546 — .0060 x	.0428	—
»	— Poids de l'utérus . . . »	—	—	55.4672 — .1477 x	.0892	—

croissance rectiligne, apparaissent quelque fois considérables. A vrai dire, la moyenne de tous les écarts carrés moyens relatifs, c'est-à-dire des écarts carrés moyens rapportés à la moyenne arithmétique, donnés dans les colonnes 3 et 5 du tableau, n'est pas trop grande. Elle est, en effet, de 4.01 % pour l'homme, et de 3.01% pour la femme. Cependant, on ne doit pas oublier que pour un caractère masculin l'écart carré moyen relatif atteint 11% et pour un caractère féminin 9.7%. Mais nous ne devons pas perdre de vue les défauts du matériel statistique dont nous nous servons, ni le rôle de simple description que nous avons attribué à l'ajustement rectiligne de l'allure de la décroissance. Comme par l'ajustement adopté nous ne prétendons pas établir une loi mais seulement décrire quelques faits, nous devons admettre que notre but est atteint le plus souvent d'une manière satisfaisante.

5. — La colonne 6 du tableau fait ressortir immédiatement que l'intensité de la décroissance est ordinairement supérieure chez la femme que chez l'homme. Sur 25 couples d'équations contenus dans le tableau, cinq fois seulement l'involution masculine apparaît supérieure à l'involution féminine (1). C'est le cas de la taille (d'après les chiffres de Bischoff) et du poids du cervelet, du poumon gauche, de l'estomac et du pancréas (d'après les chiffres de Boyd). Il suffit de quelques considérations pour réduire la portée de ces exceptions.

En ce qui concerne la taille, les équations calculées d'après les données de Powys et Pearson, Harris et Boyd, sûrement plus

---

(1) J'avais déjà livré ce mémoir à l'imprimeur, lorsque j'ai fait des calculs analogues à ceux dont il est question dans le texte, prenant pour base les données publiées par S. Weissenberg (voir ouvrage cité). J'ai trouvé que la régression sénile féminine de la taille, de la taille assise, de la grande envergure, de la longueur du tronc, des bras et des jambes, est toujours plus forte que la régression masculine correspondante; et que là où il y a progrès en fonction du vieieillissement, comme il en est le cas pour la largeur biacromiale, le diamètre par les hanches, et la circonférence par les aisselles, ce progrès est toujours plus fort chez l'homme par rapport à la femme. Voilà donc 9 couples d'équations qui ne font que confirmer (sans aucune exception) les résultats exposés dans le texte. A ceux-ci nous pouvons ajouter les 3 couples d'équations, que nous avons établis prenant pour base la taille des aliénés, et qui montrent aussi une décroissance féminine plus intense par rapport à la décroissance masculine correspondante. Nous devons avouer par là que le fait qui nous intéresse est probablement général.

dignes de foi que les chiffres de Bischoff (1) ne peuvent pas laisser de doute sur la chute plus rapide des courbes obtenues pour les femmes; et, quant au poumon gauche, nous ne devons pas oublier que le poumon droit se comporte d'une manière opposée (2).

Par conséquent, les véritables exceptions se réduisent à trois seulement, dont l'une concerne un organe (l'estomac) sur le poids duquel les processus autolytiques *post mortem*, exercent une enorme influence.

Il y a des cas où la droite d'ajustement monte au lieu de descendre. Ce fait peut être attribué soit à l'âge, quelquefois trop précoce, auquel nous avons supposé terminée la croissance (25 ans); soit au fait qu'il y a des organes, tels que le coeur, qui ne décroîtrait que fort peu, ou qui ne décroîtrait pas du tout, ou bien encore qui augmenterait jusqu'à l'extrême vieillesse (Demange, Ribbert, etc). Dans ces cas il arrive que B est positif pour l'homme et négatif pour la femme: ou, lorsqu'il est positif pour les deux sexes, sa valeur absolue est supérieure chez l'homme et inférieure chez la femme.

6. — Il n'est pas aisé de se rendre compte de la signification de ce fait singulier. Après une longue étude de la question, dont il n'est pas nécessaire de relater ici tous les détails, nous n'avons pas réussi à l'éclaircir d'une manière complète.

On peut formuler plusieurs hypothèses. Avant tout, il faut se demander si nous ne sommes pas victimes d'une illusion, étant donné que notre observation ne porte pas sur des individus qui vieillissent mais sur les survivants de chacun des groupes d'âge choisis.

Mais, en faisant abstraction de l'idée que nous discuterons plus loin — d'une sélection différente exercée par la mort parmi les individus physiquement plus ou moins développés, on ne saurait pas considérer comme accidentels des faits qui apparaissent avec une très grande constance. D'autre part, en ce qui con-

(1) J'en ai fait la critique dans ma note: *Sul peso relativo*, etc. déjà citée.

(2) Boyd, d'ailleurs, a très bien fait remarquer qu'il est rare d'observer un cadavre dont les poumons soient normaux, ce qui empêche d'attribuer trop d'importance au poid de cet organe.

La valeur très grande de B obtenue pour les poumons semble confirmer dans un certain sens ces constatations.

cerne la taille, nous sommes sûrs qu' elle décroit réellement en fonction de l' âge, et par analogie on serait tenté de dire que les poids des différents organes pourraient se comporter d' une manière analogue.

Si une régression de la taille et des poids du corps et de ses organes a réellement lieu à la suite du vieillissement, nous devons nous demander pourquoi cette régression serait plus intense chez la femme que chez l'homme.

Nous croyons avant tout devoir écarter l' idée que la différence puisse être attribuée au fait d'avoir considéré la croissance achevée au même âge chez les individus des deux sexes. En realité, lorsque les caractères quantitatifs féminins commencent à fléchir, les caractères masculins correspondants se développent encore: de là une influence sur l' allure de la courbe d' ajustement peut en resulter, courbe qui, à parité d'autres conditions, devrait descendre moins rapidement chez l'homme comparativement à la femme. Mais les valeurs absolues de B sont souvent si différentes pour les caractères correspondants des deux sexes, que cette cause d'erreur, si elle existe, est dans la plupart des cas impuissante à en donner la raison.

Le cas de la taille est très significatif à ce propos. En effet, elle montre un fléchissement plus fort chez la femme, même lorsqu' on tient compte — comme l' ont fait Powys et Pearson — de l'âge exact auquel la croissance s'achève. Il y a, d'autre part, des organes, tels que le cervau, dont le poids commence à fléchir avant même la vingtième année. Si on ne tient compte de la décroissance de ces organes qu'à partir de la 25.ème année, l'allure de la droite d' ajustement n' est pas du tout influencée par le commencement de la régression, et cette-ci est, cependant, plus intense chez la femme.

Il reste à savoir si la différence peut être attribuée à la plus grande intensité de la mortalité masculine. Si la mort choisit ses victimes parmi les individus les moins doués au point de vue physique, il est naturel que les survivants à un âge $x+h$ aient été, en moyenne, plus robustes que tous les survivants à l'âge x, les moins robustes ayant succombé dans l'intervalle h. Si on pouvait démontrer que les individus les plus développés d'une population — c'est-à-dire dont la taille et le poids du corps et des organes sont plus grands — sont aussi les plus résistants à l'ensemble des causes de décès, on comprendrait facilement que l'intensité de la régression sénile serait plus ou

moins masquée par le fait de la survivance des individus qui ont été parmi les plus développés aux âges précédents. Dans ce cas, cette erreur d'observation serait plus intense pour l'homme, soumis à une mortalité relativement plus forte, mortalité qui éliminerait dans une large mesure les individus physiquement moins développés, et moins intense pour la femme, frappée par une mortalité relativement faible. Ainsi la différence observée dans l'intensité de la régression sénile serait dûe simplement aux conditions de l'observation.

Malgré nos efforts, nous n'avons pas pu vérifier d'une manière satisfaisante l'influence du développement physique sur la mortalité. Les observations faites sur les recrues, et les assurés, qui sont contenues dans l'*Antropometria militare* italienne, et dans la *Medico-actuarial mortality investigation*, l'ouvrage americain bien connu, laisseraient croire que la mortalité des personnes de petite taille est tout au plus égale, sinon inférieure à celle de l'ensemble des recrues et des assurés. Mais pour de nombreuses raisons, que nous renonçons à développer ici, nous ne saurions pas affirmer d'une façon absolue que ce matériel, quoique très intéressant, puisse conduire à un jugement définitif (1).

(1) Une ètude ultérieure des données relatives à la mortalité des assurés américains nous a permis de constater, dès que ce mémoir avait été achevé, que le pourcentage de mortalité des jeunes hommes de petite taille et des hommes mûrs de grande taille serait inférieur au pourcentage de mortalité des jeunes hommes de grande taille et des hommes mûrs de petite taille. Par contre, le pourcentage de mortalité des jeunes femmes de petite taille et des femmes mûres de grande taille serait supérieur au pourcentage de mortalité des jeunes femmes de grande taille et des femmes mûres de petite taille. Si l'on pouvait attribuer une portée générale à ces conclusions, on devrait admettre que la décroissance sénile féminine était partiellement masquée par la tendence des femmes de grande taille à survivre, tandis que la régression sénile masculine était accentuée par la tendance à survivre des hommes de petite taille. De là, la conclusion que la différence sexuelle constatée aurait été plus marquée si on avait pu éliminer l'influence de ces facteurs.

Malheureusement, les groupes d'hommes et des femmes qui nous ont suggéré les observations précédentes n'étaient pas parfaitement homogènes entre eux, et les femmes n'étaient pas si nombreuses comme on aurait pu le désirer, de façon à nous permettre de parvenir à des resultats d'une généralité incontestable. Pour cela nous devons nous contenter d'attribuer aux faits exposés la simple valeur d'indices.

Voir notre note: *La mortalità secondo la statura nell'uomo e nella donna*. «Rassegna di Studi sessuali», 1922, pp. 161-170. Pour les données, voir: Association of Life Insurance Medical Directors and Actuarial Society of America. *Medico-Actuarial Mortality Investigation*, New York, 1912-1914.

Mais si l'on devait renoncer à expliquer la différence d'intensité de la régression sénile chez les deux sexes — telle que nous l' avons observée — par la différence de la mortalité des hommes et des femmes, il ne resterait qu'à la considérer comme un caractère sexuel secondaire. Nous tendons à admettre cette idée, quoiqu' il nous soit impossible d' en obtenir la certitude. S' il en était ainsi, nous devrions considérer la régression sénile comme un des caracterès sexuels secondaires les plus constants et les plus typiques. Mais, rien ne nous autorise à aller si loin, comme, du reste, rien ne nous donne un démenti décisif.

A part cela, nous tenons à faire remarquer le contraste existant entre l' intense régression sénile de la femme, et l' intense mortalité de l' homme. Si l' involution est un signe indubitable de déchéance physique, comment sa grande intensité pourrait-elle s' accorder avec un faible taux de mortalité?

Cependant, il n'est pas difficile de se convaincre à l'aide d' une table de mortalité d'après les causes de decès — que le caractère violent de l'homme, son intense participation aux travaux des champs, son séjour prolongé dans les milieux infects de la ville et de l' usine ont sûrement une part prépondérante dans la forte mortalité qu' il subit. Il y a lieu de se demander, alors, si nous n' assisterions pas à un excés de mortalité féminine le jour où la participation de la femme à la vie sociale et économique atteindra le même degré que chez l' autre sexe.

Si la différence de l' involution sénile de l'homme et de la femme était réellement un caractère sexuel, le doute que nous venons d' exprimer présenterait un haut degré de probabilité. Et ce serait alors le cas de se demander : quelle influence a exercé jusqu'ici sur la mortalité de la femme sa tendance progressive à se « masculiniser » ? Quels en sont le inconvénients, au point de vue de la santé individuelle et sociale, de l'hygiène et de l'eugenique? La régression sénile chez la femme en resulte-t-elle ultérieurement intensifiée, et la valeur eugénique de l' individu de plus en plus réduite ? Ou bien l'élimination des femmes moins douées au point de vue physique, ne serait-elle un resultat souhaitable conduisant à l'amélioration des races ? Voilà une foule de vieilles questions qui surgiraîent sous un nouveau jour si on pouvait démontrer fondée l'hypothèse formulée sur la regréssion sénile des deux sexes.

L' air séduisant que ces questions ont dès maintenant et les révelations qu' elles semblent nous promettre au point de vue biologique et social nous induisent à espérer de voir attiré sur elles l' attention des savants.

Le mérite de les avoir posées sous une nouvelle lumière serait, dans ce cas, le meilleur résultat de nos recherches.

Guglielmo Tagliacarne

Contributi e comportamenti delle regioni d' Italia in guerra

Sommario: **1.** Frequenza dei disertori in guerra condannati dai tribunali militari dell'esercito operante, distinguendo la regione a cui il disertore appartiene. — **2.** La frequenza dei disertori nelle varie regioni in rapporto al contributo dato alla guerra (mutilati) dalle regioni stesse. — **3.** Disertori con passaggio al nemico ed in presenza del nemico per ogni 100 disertori in complesso nelle varie regioni. — **4.** Relazione fra la frequenza dei disertori e la frequenza dei decorati nelle varie parti d'Italia. — **5.** Misura della sperequazione o differenza media fra regione e regione in guerra di fronte al contributo di sangue (mutilati), alla frequenza dei disertori e alla frequenza dei decorati.

1. — Ho potuto conoscere il numero dei disertori in guerra, distinguendo la regione in cui il disertore è nato e a seconda che questo reato di codardia si sia verificato con passaggio al nemico, in presenza del nemico o lontano dal nemico. I disertori di cui mi occupo sono soltanto quelli condannati dai tribunali militari dell'esercito operante.

Di tali disertori fu operato lo spoglio per regione, ed essi costituiscono circa il 70 per cento del totale dei soldati che disertarono.

Mancano, per ciò, i disertori condannati dai tribunali territoriali entro la zona di guerra (17 per cento) e quelli condannati dai tribunali territoriali fuori dalla zona di guerra (13 per cento).

Ciò nondimeno, i dati, riguardando una massa imponente di casi, che costituisce la grande maggioranza dei disertori, non sono certo privi di interesse.

Ho calcolato il rapporto fra il numero dei disertori e la popolazione maschile che, secondo il censimento 1911, figurava in quest'anno con un'età dai 12 ai 40 anni e che, quindi, durante la guerra costituiva la massa specifica da cui fu tratto

il nostro esercito in armi. In fine, ho ridotto a numeri indici i rapporti calcolati come sopra, considerando uguale a cento il rapporto ricavato per tutto il regno fra il numero dei disertori e la popolazione maschile sopraindicata. Ho così ottenuto i dati esposti nella tavola seguente.

TAVOLA 1

Distribuzione regionale dei disertori condannati dai tribunali militari. (1)

Regioni	Disertori con passaggio al nemico	Disertori in presenza del nemico	Disertori non in presenza del nemico	TOTALE disertori (2 + 3 + 4)	Disertori per 1000 maschi da 12 a 40 anni	Numeri indici — Base (100) il rapporto per il Regno fra disert. e pop. (9.80)
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Piemonte . . .	249	533	4.295	5.077	6.94	*71*
Liguria	53	134	1.192	1.379	4.95	*50*
Lombardia . .	237	927	9.973	11.137	10.70	*109*
Veneto	534	1.161	8.779	10.474	15.23	*155*
Emilia	139	387	3.770	4.296	7.58	*77*
Toscana . . .	134	355	3.169	3.658	6.44	*66*
Marche	35	102	746	883	4.29	*44*
Umbria. . . .	28	97	686	811	5.91	*60*
Lazio	34	245	3.894	4.173	14.24	*145*
Abruzzi e Molise	58	179	1.238	1.475	6.36	*65*
Campania . . .	104	758	10.521	11.383	17.71	*181*
Puglie	85	377	4.076	4.538	10.20	*104*
Basilicata . . .	21	76	723	820	9.44	*96*
Calabrie . . .	44	241	2.145	2.430	9.00	*101*
Sicilia	159	592	5.947	6.698	8.46	*86*
Sardegna . . .	16	51	627	694	3.73	*38*
Estero	27	71	627	725	—	—
Ignoti	41	29	121	191	—	—
Totale	1.998	6.315	62.529	70.842	9.80	*100*
Italia settentr. .	1.212	3.142	28.009	32.363	9.79	*100*
» centrale .	289	978	9.733	11.000	7.66	*78*
» meridion.	254	1.452	17.465	19.171	13.50	*138*
» insulare .	175	643	6.574	7.392	7.56	*77*

(1) Nel 1921 il MINISTERO DELLA GUERRA (DIVISIONE GIUSTIZIA MILITARE) pubblicò una relazione dal titolo *La delinquenza militare durante il periodo bellico.* In tale relazione figurano le cifre complessive per tutta l'Italia

Nella tavola precedente la Sardegna appare con la quota minima (numero indice 38) subito seguita dalle Marche (num. ind. 44); invece le quote più alte sono rappresentate dalla Campania (num. ind. 181), a cui tengono dietro il Veneto (num. ind. 155) e il Lazio (num. ind. 145).

2. — Ma è chiaro che per avere un indice che riveli in modo più significativo il comportamento delle varie regioni di fronte al reato di diserzione si dovrebbe tener conto del diverso contributo che le varie regioni hanno dato alla guerra.

Ad esempio, il basso numero indice che nella nostra tavola figura per la Liguria si deve mettere in relazione con la scarsa partecipazione di questa regione alla logorante guerra di trincea, a causa dei numerosi esoneri concessi agli operai delle industrie e al largo contributo che questa regione ha fornito alle marine di guerra e mercantile.

Epperò per poter misurare il contributo dato alla guerra dalle varie regioni, mancano sia il numero dei militari forniti all' esercito dalle diverse parti d' Italia, sia il numero dei morti che ogni regione ha dato per la causa comune. Solo si hanno dei dati circa la distribuzione regionale di 30.770 mutilati. Si tratta di una pubblicazione della Sanità militare (1), la quale anticipa i risultati di una parte dello spoglio operato sulle schede dei mutilati. Le schede per ora spogliate si riferiscono ai mutilati il cui cognome incomincia per le lettere A, B, C, D,

---

degli imputati e dei condannati militari per i vari reati. Le cifre che si riferiscono alla diserzione, secondo la pubblicazione citata, sono le seguenti (riporto solo le notizie per i condannati):

Disertori con passaggio al nemico	2.022
» in presenza del »	6.335
» non in presenza del »	93.308
Totale disertori	101.665

70.707 disertori furono condannati dai tribunali di guerra dell'esercito operante, 17.474 dai tribunali territoriali entro la zona di guerra e 13.484 dai tribunali territoriali fuori della zona di guerra.

Si noterà che la cifra complessiva indicata nella nostra tavola 1 per i disertori condannati dai tribunali di guerra (70.842) non coincide perfettamente con quella fornita dalla relazione del Ministero, ora citata (70.707).

La differenza è però trascurabile e si deve imputare (almeno, così mi hanno informato) al fatto che dopo che fu pubblicata la surriferita relazione del Ministero della guerra, furono rintracciate un centinaio di altre schede, che si fecero figurare nelle classifiche per regioni, riportate nella nostra tavola 1.

(1) Ministero della guerra (Direzione generale di sanità militare): *Dati statistici su 30.770 invalidi di guerra.*

E, F. — I dati forniti dalla pubblicazione sopra ricordata sono in cifre assolute e riguardano le varie provincie; io mi limito a indicare i dati per regione e calcolo quindi il rapporto fra mutilati e popolazione maschile da 12 a 40 anni (censim. 1911), presentando nella colonna (4) i numeri indici che si ottengono ponendo uguale a cento il rapporto ricavato per il regno (4.29).

Infine, nell' ultima colonna, faccio il rapporto fra i numeri indici dei disertori (col. (7) della tavola 1) e i numeri indici dei mutilati (col. (4) della tavola seguente) e moltiplico il quoziente per cento. Gli indici così ottenuti possono forse mostrare, meglio di quelli presentati nella prima tavola, il contegno delle varie regioni per rispetto al reato di diserzione.

Ora la quota minima viene ad essere rappresentata dalle Marche (num. ind. 40), seguite dagli Abruzzi (num. ind. 44) e dalla Sardegna (num. ind. 45); la Campania figura col numero indice più elevato (202) e viene appresso il Lazio (num. ind. 169).

3. — Si osservi, però, ad attenuare l' impressione che potrebbero lasciare queste ultime cifre, come esse cifre risultino nella loro composizione.

Infatti, mentre, per il regno, fra cento disertori se ne contano 11.73 con carattere di maggiore gravità (passaggio al nemico e in presenza del nemico), se ne hanno soltanto (veggasi tavola N. 3) 7.57 nella Campania e 6.69 nel Lazio, che sono, come si è visto poc' anzi, appunto le regioni in cui si è constatato la maggiore frequenza dei disertori.

Anche per rispetto alle varie zone d' Italia si può fare la stessa osservazione. Mentre considerando tutte insieme le varie forme di diserzione, l' Italia meridionale (in ragione della sua partecipazione alla guerra) figura con la maggior quota di disertori, quando invece si osservi la composizione della massa dei disertori, si rileva che appunto l' Italia meridionale ci dà, in proporzione, il minor numero dei casi in cui il reato di diserzione si presenta nelle forme più gravi. Per cui l' impressione che si riceve sulle prime, osservando la maggiore diffusione di questo reato in certe regioni, si viene a correggere quando si ponga mente alla gravità del reato stesso, com' è posto in luce dalla colonna dei numeri indici della frequenza dei disertori con passaggio al nemico e di fronte al nemico, ragguagliata alla popolazione.

TAVOLA 2

Distribuzione regionale di 30.770 mutilati e misura della frequenza dei disertori ragguagliata alla frequenza dei mutilati.

Regioni	Mutilati			N. I. disertori × 100 / N. I. mutilati
	Cifre assolute	Per 1000 maschi da 12 a 40 anni	Num. indici Base (100) il rapporto per il regno fra mutilati e pop. mas. 12-40 anni (4.29).	
(1)	(2)	(3)	(4)	(5)
Piemonte	2849	3.89	*91*	*78*
Liguria	664	2.38	*55*	*90*
Lombardia	3737	3.59	*84*	*130*
Veneto	3470	5.04	*117*	*132*
Emilia	2394	4.22	*98*	*79*
Toscana	2593	5.00	*117*	*56*
Marche	959	4.65	*108*	*40*
Umbria	644	4.69	*109*	*55*
Lazio	1084	3.69	*86*	*169*
Abruzzi e Molise	1457	6.28	*146*	*44*
Campania	2472	3.84	*89*	*202*
Puglie	1723	3.87	*90*	*115*
Basilicata	345	3.97	*93*	*104*
Calabria	1072	4.36	*102*	*99*
Sicilia	3024	3.82	*89*	*97*
Sardegna	680	3.65	*85*	*45*
Estero	136	—	—	—
Non indicata	1467	—	—	—
Totale	30770	4.29	*100*	*100*
Italia settentrionale	13.114	3.97	*93*	*108*
» centrale	6.737	4 69	*109*	*71*
» meridionale	5.612	3.95	*92*	*150*
« insulare	3.704	3.79	*88*	*87*

Questa serie di numeri indici ha un'importanza particolare. Essa considera il fenomeno nella sua espressione più morbosa e più significativa; d'altra parte, essa tiene conto di quasi tutti i casi consimili verificatisi, come si rileva confrontando le cifre complessive delle colonne (2) e (3) della tavola 1 con le analoghe cifre riportate nella nota alla stessa tavola 1. Infatti i disertori con passaggio al nemico e in presenza del nemico ammontano, secondo la nostra tavola 1, complessivamente a 8.313, contro 8.357 che figurano nella relazione del Ministero della guerra e che rappresentano il totale di tutti i disertori che commisero tale reato di codardia nelle condizioni sopra ricordate (presenza del nemico e passaggio al nemico). Invece i disertori che disertarono lontani dal nemico sono in complesso (relazione del Ministero della guerra) 93.308, contro 62.529 considerati nella nostra tavola 1.

Può quindi sorgere legittimo il dubbio che la ripartizione nelle varie regioni degli altri 30.779 militari da noi trascurati, che disertarono lontani dal nemico e che saranno stati condannati dai tribunali territoriali entro la zona di guerra e dai tribunali territoriali fuori dalla zona di guerra possa modificare quella graduatoria risultata dalla serie di numeri indici della nostra tav. 1, mentre siffatto dubbio non può nascere quando si considerino i soli disertori con passaggio al nemico e in presenza del nemico, come si fa con la serie di numeri indici della tavola seguente.

Non si può a questo punto tralasciare di osservare la particolare posizione che viene ad assumere la Sardegna.

Questa regione che ha di già uno scarso numero di diserzioni nel complesso, appare con un numero veramente esiguo di disertori che passano al nemico o disertano in presenza di esso.

Quantunque in questo breve saggio non vi sia neppure il tentativo di stabilire una graduatoria fra le varie regioni per riguardo al comportamento di esse di fronte al reato di diserzione nella recente guerra, mancando troppi elementi indispensabili per un ponderato giudizio, pure si deve mettere in rilievo il fatto che i dati sin qui esposti concordano nell'attribuire alla Sardegna il posto d'onore.

Il forte numero dei disertori del Veneto si deve spiegare in gran parte con la circostanza che, durante la ritirata di Caporetto, i veneti avevano una particolare tentazione di fermarsi in paese. Del resto vedremo fra poco che i veneti hanno ripor-

TAVOLA 3

Distribuzione regionale dei disertori con passaggio al nemico o in presenza del nemico.

Regioni	Disertori con passaggio al nemico e in presenza del nemico			Numeri indici
	Cifre assolute	Per ogni 1000 maschi dai 12 ai 40 anni	Su 100 disertori in complesso	Base (100) il rapporto per il Regno fra disertori con passaggio al nemico e in presenza del nemico e popolazione maschile dai 12 ai 40 anni (11.40)
(1)	(2)	(3)	(4)	(5)
Piemonte	782	10.70	15.40	*94*
Liguria	187	6.71	13.56	*59*
Lombardia	1164	11.19	10.45	*98*
Veneto	1695	24.65	18.18	*216*
Emilia	526	9.28	12.24	*81*
Toscana	489	8.61	13.37	*75*
Marche	137	6.66	15.52	*58*
Umbria	125	9.10	15.41	*80*
Lazio	279	9.52	6.69	*83*
Abruzzi e Molise	237	10.22	16.07	*90*
Campania	862	13.41	7.57	*118*
Puglie	462	10.39	11.18	*91*
Basilicata	97	11.17	11.82	*98*
Calabrie	285	11.61	11.73	*102*
Sicilia	751	9.49	11.21	*83*
Sardegna	67	3.60	9.65	*32*
Regno	8145	11.41	11.73	*100*
Italia settentrionale	4354	13.17	13.45	*115*
» centrale	1267	8.82	11.52	*77*
» meridionale	1706	12.02	7.90	*105*
» insulare	818	8.37	11.07	*73*

tato gran numero di medaglie al valore. L'alto numero di disertori del Veneto concorre a mantenere alquanto elevato anche l' indice dell' Italia settentrionale.

4. — Ho voluto anche tentare di scoprire se vi sia relazione fra il numero dei disertori ed il numero dei decorati al valor militare delle varie regioni.

Per quanto riguarda i decorati mi riferisco a uno spoglio da me compiuto sopra 80.942 medaglie al valor militare concesse dall'inizio della guerra sino al 15 luglio 1920.
Le medaglie al valore sono d'oro, d'argento e di bronzo. Ho creduto opportuno di ridurle tutte allo stesso valore, cioè tutte a medaglie di bronzo, trasformando il carattere qualitativo della medaglia in un carattere quantitativo in base all'assegno annuo, che è stabilito in L. 800 per le medaglie d'oro, L. 250 per quelle d'argento e L. 100 per quelle di bronzo.

In tal modo si tiene conto non solo della *diffusione*, ma anche dell'*intensità* del valore dimostrato dalle diverse regioni.

Si deve, però, avvertire che i numeri indici che si presentano nella tavola seguente riguardo al numero di medaglie concesse in ogni regione in rapporto alla popolazione maschile da 12 a 40 anni (censimento 1911) non possono costituire la base per una graduatoria di valor militare fra le varie regioni.

Mancano, per poter stabilire una simile graduatoria, analogamente a quanto si è già detto esaminando i dati dei disertori, parecchi elementi, che non si dovrebbero trascurare per una conclusione di tale portata.

Così, se la Liguria nella tavola seguente figura in uno degli ultimi posti, non si può dire per ciò solo che i liguri siano stati, per quanto a valore in guerra, da meno degli altri italiani, giacchè, come ho avuto occasione di ricordare più sopra, la Liguria ha dato alla guerra un contributo minore delle altre regioni (si veda la quota di mutilati, tavola n. 2), per cui era quivi d' attendere un minor numero di decorati, il quale pertanto, non può costituire una prova di inferiorità nel valore. Chè anzi, se si tien conto dello scarso contributo fornito dalla Liguria all'esercito combattente, vediamo i liguri apparire fra i più ardimentosi (veggasi tav. N. 5); liguri, in altre parole, sarebbero stati pochi al fronte, ma quei pochi si sarebbero distinti per valore, come si rileva dallo specchietto seguente, nel quale si sono appunto calcolati i rapporti fra decorati e mutilati.

TAVOLA 4

La distribuzione regionale di 80.942 medaglie al valore militare.

Regioni	med. d'oro	med. d'argento	med. di bronzo	Totale medaglie	Totale medaglie ridotte a bronzo	Medaglie ridotte a bronzo per ogni 1000 maschi da 12 a 40 anni	Numeri indici — Base (100) il rapporto del regno fra decorati e popolazione maschile da 12 a 40 anni
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Piemonte . . .	37	3983	5848	9868	16101	220.23	*124*
Liguria	8	979	1444	2431	3955	141.91	*80*
Lombardia . . .	31	4576	7799	12406	19487	187.26	*106*
Veneto	23	3362	5666	9051	14255	206.90	*117*
Emilia	18	2423	3834	6275	10036	177.02	*100*
Toscana	19	2284	3783	6086	9645	169.80	*96*
Marche	6	865	1483	2354	3694	179.44	*101*
Umbria	1	604	998	1603	2516	183.23	*104*
Lazio	9	1546	2162	3717	6099	208.14	*118*
Abruzzi e Molise	11	1097	1648	2756	4478	193.08	*109*
Campania . . .	17	2329	3567	5913	9526	148.22	*84*
Puglie	4	1452	2258	3714	5920	133.09	*75*
Basilicata . . .	3	312	497	812	1301	149.76	*85*
Calabrie	9	890	1435	2334	3732	152.04	*86*
Sicilia	16	2420	3696	6132	9874	124.73	*70*
Sardegna . . .	13	1066	1424	2503	4193	225.37	*127*
Estero	9	791	1234	2034	3284	—	—
Non indicata . .	—	393	560	953	1542	—	—
Regno	234	31372	49336	80942	129638	177.00	*100*
Italia settent. . .	117	15323	24591	40031	63834	193.15	*109*
» centrale .	46	6396	10074	16516	26432	184.05	*104*
» merid. . .	33	4983	7757	12773	20479	144.24	*81*
» insulare .	29	3486	5120	8635	14067	143.89	*81*

TAVOLA 5

Regioni	N. ind. decorati × 100 / N. ind. mutilati
Piemonte	137
Liguria	144
Lombardia	126
Veneto	99
Emilia	102
Toscana	82
Marche	94
Umbria	95
Lazio	137
Abruzzi e Molise	75
Campania	94
Puglie	83
Basilicata	91
Calabrie	85
Sicilia	79
Sardegna	150
Regno	100
Italia settentrionale	117
» centrale	95
» meridionale	88
» insulare	92

Nello specchietto precedente, per meglio valutare il carattere ardimentoso delle varie regioni, si è inteso di tener conto del diverso contributo che le differenti parti d'Italia hanno avuto occasione di dare alla guerra, e si è assunto come misura di tale contributo, in mancanza di dati più diretti, il numero dei mutilati. Si tenta in tal modo di scoprire quale sia stata la frequenza dei decorati fra quei tali piemontesi, quei liguri, quei lombardi, ecc., che presero parte attiva alle operazioni belliche.

Ma errerebbe anche chi credesse di poter assumere le cifre così ottenute per istabilire una graduatoria di merito, di valo-

rismo fra regione e regione, chè la frequenza regionale dei decorati non è sempre in funzione del maggior valore di cui furono animati i combattenti delle varie parti d'Italia.

E' notorio, per esempio, che gli ufficiali e i graduati ottennero, in proporzione al loro numero, assai più medaglie della truppa. Ciò si deve in gran parte a una maggiore distinzione, anche per quanto a valore, degli ufficiali e dei graduati. Spesso, anzi, il grado era conseguito appunto in seguito a prove di valore. In altre parole, ufficiali e graduati costituivano già una massa selezionata e il criterio della scelta fu spesso il valore dimostrato in guerra.

Ma non si può tacere che la maggiore frequenza delle medaglie agli ufficiali e ai graduati si deve ricercare anche nel fatto che essi erano più in vista e, rappresentando il reparto, spesso la medaglia era a loro concessa più facilmente o come segno di riconoscimento del valore dimostrato dal reparto.

Ne deriva che le regioni le quali maggiormente fornirono all'esercito ufficiali e graduati (sono specialmente le regioni dell' Italia settentrionale) figurino con un maggior numero di decorati, che non sempre corrisponde a un maggior numero di valorosi.

Limitando, come ho fatto, la portata delle cifre ricavate dallo spoglio da me compiuto, non voglio dire che tali cifre siano prive di interesse. Esse hanno pur sempre un notevole valore indiziario e, non fosse altro, valgono come sicura riprova della conclusione a cui sopra sono pervenuto nei riguardi della Sardegna.

Infatti dai dati dei due precedenti specchietti relativi alla frequenza dei decorati, vediamo che la Sardegna compare sempre al primo posto, confermando in tal modo il suo tradizionale spirito ardimentoso, per cui i sardi si distinsero tanto nobilmente in guerra. Così mentre da un lato abbiamo veduto questa regione apparire ultima per riguardo al numero dei disertori, la vediamo, ora, prima nella gara di valore.

E, quindi, per mezzo di due caratteri opposti si prova in modo sicuro quale sia stato il contegno dei sardi durante la guerra.

Ho misurato quale relazione esista fra la graduatoria delle varie regioni per rispetto alla frequenza dei disertori (colonna (7) della tavola N. 1) e la graduatoria secondo la frequenza di medaglie al valor militare (colonna (8) della tavola N. 4), adottando all'uopo la formula a cui é pervenuto il Gini per gli

indici di concordanza tra le graduatorie. (1) L' indice ottenuto è — 0,11.

Sostituendo nel calcolo, alla frequenza dei disertori in complesso quella dei soli disertori con passaggio al nemico e in presenza del nemico (colonna (5) della tavola N. 3) l' indice risulta — 0,08.

Perciò possiamo dire che fra i due caratteri esaminati, frequenza dei disertori e frequenza dei decorati nelle varie regioni, c'è l'accenno a una controgradazione, per cui le regioni che danno il minor numero di disertori tendono a presentarsi col maggior numero di decorati; però la controgradazione riscontrata è debole.

5. — Ho voluto, infine, misurare la sperequazione in guerra, fra regione e regione, di fronte al contributo di sangue (frequenza dei mutilati), alla frequenza dei disertori e alla frequenza dei decorati, per tentare di scoprire quale sia stata in quest'occasione l'importanza della disparità di sforzi e di comportamento delle diverse regioni italiane e in quale delle tre forme esaminate la differenza fra una regione e l'altra sia stata più grave.

Ho quindi calcolato degli indici di variabilitá, o in altre parole, di « quanto le diverse quantità rilevate differiscano fra di loro », adottando come indice di variabilità la differenza media Δ (2).

TAVOLA 6

	Differenze medie fra regione e regione	Δ
A)	dei numeri indici della frequenza di medaglie al val. mil. (col.8, tav. 4)	20,83
B)	„ „ „ „ „ „ mutilati (col. 4, tav. 2) . . .	21,48
C)	„ „ „ „ „ „ disertori, in complesso (col. 7, tav. 1)	46,72
D)	„ „ „ „ „ „ „ con passaggio al nemico e in presenza del nemico (col. 5, tav. 3)	37.05

Gli indici presentati in questo specchietto ci avvertono che la sperequazione o differenza media fra regione e regione è minima quando si consideri il numero di medaglie concesse al

(1) CORRADO GINI: *Di una misura delle relazioni tra le graduatorie di due caratteri.* — Roma, 1914, — Tipografia ditta Ludovico Cecchini.

(2) CORRADO GINI: *Variabilità e mutabilità* — Estratto dagli « Studi economici-giuridici » pubblicati a cura della facoltà di giurisprudenza della R. Università di Cagliari — Anno III, parte 2ª.

valor militare e conserva quasi la stessa intensità per rispetto al contributo di sangue, ma si intensifica notevolmente di fronte al reato di diserzione.

Epperò per rispetto a tale reato la sperequazione appare minore quando si esamini il fenomeno nella sua manifestazione più grave (passaggio al nemico e in presenza del nemico). Ciò ha particolare importanza per il fatto già accennato che quasi tutti i casi di questa diserzione più grave e più significativa sono considerati nelle nostre statistiche.

E si aggiunga che la differenza media (Δ) per tale diserzione (lettera D), essendo calcolata su un complesso di elementi notevolmente più esiguo in confronto alla massa totale dei disertori (lettera C) dovrebbe perciò essere, a parità di ogni altra circostanza, assai più rilevante e quindi la diminuzione che invece si riscontra fra le differenze medie C) e D) viene ad acquistare maggiore importanza e un più preciso significato.

Pertanto, ove per rispetto alla diserzione si tenga conto soltanto di quella che presenta caratteri di maggiore gravità, com'è consigliabile, e d'altra parte si ponga mente al diverso numero di casi compresi nel calcolo delle varie differenze medie indicate nello specchietto precedente (lettere (A, B, D), possiamo infine concludere che la sperequazione fra regione e regione non pare sia molto dissimile a seconda che le varie regioni siano considerate di fronte ai sacrifici di sangue ovvero alla frequenza dei decorati o ancora alla frequenza dei disertori.

ARTHUR MACDONALD

Death Psychology of Historical Personages

A summary of the last words of those distinguished people in history, where records have come down to us, is a psychology of their death.

Before presenting the results of such a summary, and in order to understand better the significance of words during the dying hour, it may be useful to note a few points as to death itself.

The Dying Hour.

In another place, the author has treated in detail the physiology and psycology of death (1). Here it may be stated that death is neither rapid nor sudden, but is preceded by a period of transition, which begins as soon as the reactionary forces of the organism have ceased and combat has ended.

The death act is often confounded with the symptoms of disease which precede it. Dying begins after these symptoms have subsided, there is a pause in nature, the disease has conquered, the battle is over, and all is tranquil.

This transition stage, or dying hour, may last for a longer or shorter time; in the great majority of cases persons are unconscious, thus the natural death appears to be a brain death.

But when there is consciousness during the hour, it depends upon nutrition and provision of the brain with blood. As there are three ways of physical death (1) by brain by heart and by lungs, so there are three kinds of psycological death. The first is little or no delirium, and intelligence continues not only to the end but

(1) See *Anthropology of modern civilised Man* published by the *Anthropological Society* of Bombay; also in «Western Medical Times». Denver. Colorado. July, 1920; also in «Medical Times» N. Y. City 1920. — See also « Medical Times ». N. Y. City, June 1921, and «Western Medical Times» Denver, Colorado, July 1921.

becomes very acute; physical prostration appears to be replaced by intellectual exaltation. Another kind of psychological death refers to diseases only secondarily connected with the brain; the mind is a mixed state between reason and delirium. The third kind of mental death includes all the lesions of the brain, which are almost always accompanied with loss of understanding; delirium is a symptom, there is a general obscuration of intelligence, and complete loss of consciousness.

Fear of Death

In life the fear of anyting is often much worse than the thing itself. This is especially true in the case of death. When the dying hour comes, the fear of death disappears.

Whether it is the brain, heart or lungs which give the signal of death, the brain forces are usually weakened or destroyed first, causing sensation to lessen or cease. Wether there be consciousness to the last, or at times, depends upon the nature of the disease and the mental and moral character of the person dying: and this in connection with surrounding conditions. In old age death is the last sleep, showing no difference from normal sleep. The general consensus of opinion based upon the experience of all ages is, that the dreadfulness of death and its physical pain are for the most part in the imagination.

Psychological Summary of Death of Distinguished Persons in History.

The average man usually dies unknown, whatever he thinks and says is soon forgotten. Fine and significant words may be lost. Now and then a physician may take the pains to note the last words of some of his patients. It may be a dozen cases, or more, but it is small compared with the great number of those dying every day. Therefore almost all the last words recorded at death are those of eminent and distinguished persons.

This table represents a first attempt to summarize the mental condition, at or just before death, of distinguished persons from the beginning of history up to the present time. Only the most reliable sources have been utilized, and even here, where there appeared to be any doubt the persons were omitted, so that we have remaining but 894 cases. It must be remembered howewer

that very few death-bed experiences are published, and still fewer described with sufficient accuracy to be made the object of scientific study.

Taking in to consideration the very many and varied sources, in all periods of history, which have been consulted, the regularity of the figures in the table is remarkable. Such uniformity coming out of most heterogeneous conditions, when put into statistical form, suggests that death is a great equalizer and leveler for all humanity. In a way, the conduct and last words of those facing death, are a mental and moral test of their real character.

The persons, whose records we have studied, are classified according to occupation, into ten divisions, as indicated in the first column of the table. Those whose profession was of a religious character, are the largest in number (192), which is due doubtless to the power of religion throughout all history. Under « philosophers » are included mathematicians and educators. As the number of women was not large enough to make subdivisions, they are all placed together.

It will be seen from the third column of the table, giving the average age, that the great majority of men who become eminent must live at least fifty years. The Royalty and military show the lowest average age, due in part to the large number of deaths by violence, which is the case also with the Religious, Statesmen and women. In short, all the ages in divisions where there are many deaths by violence, would of course have a much higher average age had they lived their natural lives.

Eliminating this factor of death by violence, the poets and artists die the youngest, thus Keats died at 26, Byron at 36, Burns 37, Poe 38, and Addison at 47. In the columns for pain or little or no pain at death, it will be seen that in only 80 cases out of 894, was any reference made to this matter, indicating that the question of pain at death is regarded as of little importance. In the last 14 columns of the table is presented the mental state at death or just before death, as shown by the last words. It will be noted from column 13 at the bottom, that 17 percent were sarcastic or jocose, indicating a high degree of mental control. In fact some of the dying complained that it was taking too long and they were getting tired. A relatively large number (24, or 37 percent) of writers and authors (literateurs) were jocose or sarcastic, or both (column 13); they also were relatively the freest from pain (column 10).

Psychology of death of disti

PROFESSION or OCCUPATION	Manner of Death								Number
	Number of persons	Average Age	Violence					Disease	
			Executed	Killed in battle	Suicide	Assassinated	Total		Little or no Pain
1	2	3	4	5	6	7	8	9	10
Religious	192	56	42	0	0	4	46	146	12
Royalty	91	50	23	5	2	7	37	54	5
Military	75	51	21	21	1	6	49	26	3
Philosophers	102	66	6	0	5	2	14	88	10
Literateurs	106	64	7	2	1	2	11	95	5
Physicians and Scientists . .	54	64	3	0	0	0	3	51	4
Artists	39	63	0	0	0	0	0	39	4
Poets	69	56	3	1	1	0	5	64	3
Statesmen	96	65	18	1	2	6	27	69	6
Women	70	56	19	0	3	0	22	48	7
Totals or averages . . .	894	60	142	30	15	27	214	680	59
Percentages	—	—	16	3	2	3	24	76	65

guished persons in history

Mental State of Dying Persons at Death													
Last Words at or just before Death													
Number of Persons							Average Number of Words						
Religious	Sarcastic, Jocose	Requests, directions, admonitions	Question, answer, exclamation	Contented	Discontented	More or less indifferent	Religious	Sarcastic, Jocose	Requests, directions, admonitions	Questions, answer, exclamations	Contented	Discontented	More or less indifferent
12	13	14	15	16	17	18	19	20	21	22	23	24	25
39	6	26	10	66	17	30	23	10	17	2	23	22	19
11	7	8	11	13	15	16	18	17	23	6	20	23	12
2	1	20	8	15	12	17	11	0	16	6	11	16	17
6	6	11	14	28	8	17	9	18	11	5	8	11	20
9	24	8	9	23	10	27	18	12	14	5	14	15	16
4	1	8	3	15	3	10	9	0	8	5	22	11	10
3	2	0	8	7	4	9	21	13	0	5	12	16	9
4	5	8	8	17	2	18	20	24	14	3	15	16	15
8	5	14	10	26	14	17	14	45	18	3	10	14	12
7	3	9	6	17	11	9	12	34	63	8	11	15	15
93	60	112	87	227	96	170	18	20	19	5	16	16	14
26	17	31	26	46	19	35							

The military show much the relatively highest number of requests, directions or admonitions (column 14) in their last words. The philosophers stand relatively high in questions, answers, and exclamations (column 15). In general it will be noted (columns 12-15) that requests, directions, and admonitions were most frequent (31 per cent).

More than twice as many (46 percent) were contented than discontented (19 per cent), as seen at end of columns 16 and 17; this accords with the fact that 65 percent had little or no pain and 35 percent pain. Thirty- five percent were indifferent (column 18), but they all took about the same number of words to express their feelings (averages columns 23, 24, 25). While relatively few of the Statesmen and women were sarcastic (column 13), they took many more words to express it (column 20), than the others, the poets also had as high an average as 24 words. In requests, di rections, and admonitions the women show an average of 63 words, which is three times as great as any of the others, except Royalty which was 23 (column 21).

As to expressing contentment or discontentment, the religious and Royalty used the most words, except for contentment the physicians and scientists had an average of 22 words (column 23). The artists and scientists used the fewest words of all (averages 9, 10), to express their indifference (column 25).

References

Book of Death.

Chambers *Biographical Dictionary.*

Egbert Walter R. *Last Words of Famous Men and Women,* Norristown, Pa., 1898, 792 pages.

Kaines Joseph. *Last Words of Eminent Persons,* London, 1866, 397 pages.

Lippincott, *Biographical History.*

Marvin Frederick R. *The Last Words of Distinguished Men and Women,* 1901, 80, 336 pages.

A. Andréadès

Prof. ord. de Sciences de Finances à l'Université d'Athènes

Histoire de la dette publique grecque (1e partie, Athènes, 1904) . . . . 8 francs

Les postes grecques avec l'introduction des timbres-postes (Athènes, 1906) . . . . 3 francs

Les dépenses constitutionnelles (Athènes, 1907) . . 3 francs

Du recensement (Athènes, 1908) . . . . 5 francs

Les interventions financières internationales et la doctrine de Drago (Athènes, 1910) . . . . 5 francs

Gladstone financier (Athènes, 1911) . . . . 5 francs

L'impôt sur les successions (Athènes, 1913) . . . 5 francs

L'administration économique et financière des Venitiens dans les Iles Ioniennes (2 volumes, Athènes, 1914) . . . 30 francs

Les finances publiques de Sparte (Athènes, 1915) 6 francs

Les théories financières d'Aristote et de l'école péripatéticienne (Athènes, 1915) . . . . 6 francs

Des monnaies byzantines et du pouvoir d'achat des métaux précieux dans l'empire d'Orient (Athènes, 1918) . . 6 francs

L'émigration grecque aux États-Unis (Athènes, 1917) 20 francs

Histoire des finances grecques depuis l'antiquité jusqu'à nos jours (tome 1er, pages 636, Athènes, 1918) . . . . . 30 francs

Tous ces ouvrages sont en vente chez la librairie Elefteroudakis, Place de la Constitution, Athènes.

E. Czuber

o. Prof. an der Technischen Hochschule in Wien

Die statistischen Forschungs-methoden — Verlag von L. W. Seidel & Sohn, Wien, 1921, X u. 238 Seiten . . . .

M. Greenwood and E. L. Collis

The Health of the Industrial Worker — London, 1921, I. and A. Churchill, ... Price 30 sh.

Friedrich von Fellner

o. öff. Universitäts Professor in Budapest

Das System der Rentengüter und seine Anwendung in Ungarn — Von der Ungarischen Akademie der Wissenschaften Preisgekrönt — Puttkammer und Mühlbrecht, Berlin 1905. Preis M. 120.

Die Zahlungsbilanz Ungarns — Ein Beitrag zur Lehre von der Internationalen Zahlungsbilanz im Allgemeinen. Wiener Staatswissenschaftliche Studien — Franz Deuticke, Wien und Leipzig-1908 . . . . Preis M. 105,60

Die Währungsreform in Ungarn mit besonderer Rücksicht auf die Aufnahme der Barzahlungen — Im Auftrag der Ungarischen Akademie der Wissenschaften — Manz'sche k. u. k. Hof-Verlags- und Universitätsbuchhandlung, Wien 1911 . Preis M. 90

Das Volkseinkommen Oesterreichs und Ungarns — Manz'sche k. u. k. Hof- Verlags- und Universitätsbuchhandlung, Wien 1917 Preis M. 75

Corrado Gini

Prof. ord. di Statistica nella R. Università di Padova

Il sesso dal punto di vista statistico. Palermo, Remo Sandron, 1908 . . . . Lire 8

I fattori demografici dell' evoluzione delle nazioni. Torino, Bocca, 1912 . . . . Lire 6

L' ammontare e la composizione della ricchezza delle nazioni. Torino, Bocca, 1914 . . . . Lire 13

Problemi sociologici della guerra, Bologna, Zanichelli, 1921. Lire 32

Armand Julin

Secrétaire général du Ministère de l'Industrie et du Travail de la Belgique

Principes de statistique théorique et appliquée. Premier volume: *Statistique théorique.* Bruxelles, Albert Dewit, 1921 . . 45 Frs.

Lucien March

Directeur de la Statistique générale de la France

La grammaire de la science par Karl Pearson, traduction en français par L. March — Editeur Alcan & Lisbonne, Paris. . . . .

Raymond Pearl

Professor of Vital Statistics and Biometry in the John Hopkins University, Baltimore

Modes of Research in Genetics — Published by the MacMillan Company, New York. . . . . Price $ 1,75

The Nation's Food — Published by W. B. Saunders Company, Philadelphia . . . . Price $ 3,50

The Biology of Death — Published by the J. B. Lippincott Company, Philadelphia . . . . Price $ 2.50